```
JOURNAL    Mol. Microbiol. 14 (2), 217-233 (1994)
MEDLINE    95131744
PUBMED     7830568
REFERENCE  2 (bases 1 to 4317)
AUTHORS    St. Geme,J.W.
TITLE      Direct Submission
JOURNAL    Submitted (17-JUN-1994) Joseph W. St. Geme, Molecular Microbiology,
           Washington University School of Medicine, 660 South Euclid Ave.,
           St. Louis, MO 63110, USA
COMMENT    On Jul 12, 1996 this sequence version replaced gi:562013.
FEATURES            Location/Qualifiers
     source         1..4317
                    /organism="Haemophilus influenzae"
                    /strain="N187"
                    /db_xref="taxon:727"
     gene           61..4242
                    /gene="hap"
     CDS            61..4242
                    /gene="hap"
                    /standard_name="Haemophilus adhesion and penetration"
                    /note="HAP"
                    /codon_start=1
                    /transl_table=11
                    /product="adhesion and penetration protein"
                    /protein_id="AAB03707.1"
                    /db_xref="GI:1408569"
                    /translation="MKKTVFRLNFILACISLGIVSQAWAGHTYEGIDVQYYRDFAENK
                    GKFTVGAQNIKYVNKQQLVGTSMTRAPMIDFSVVSRNGVAALVENQYIVSVARNVGY
                    TDVDFGAEGNNPDQHRFTYKIVKRNNYKKDNLHPYEDDYHNPRLHKFVTEEAPIDMTS
                    NWNGSTYSDRTKPYPERVRIGGSRQFWRNDQDKGDVAGAYHYLTAGNTHNORGAGNGY
                    SYLGGDVRKAGEYGPLPIAGSKGDSGSPMFIYDAEKQXWLINGILREGNPFEGKENGF
                    QLVRKSFPDEIFERDLHTSLSTIRAGNGVYYISGNDNGOOSITOKRSGIPSEIKITLANM
                    SLPLKEKDKVHNPRDYGPNIYSPRLNNGETLYFMPDOXQGSLIFASDINQGAGLIYFEG
                    NFTVSPNSNQTWQGAGIHVSNSNSTVWKVNGVEHDRLSKIGKGTLHVQAKGNKGSIS
                    VGDCKVILEOQADDQGNKQAFSEIGLVSGRGTVQLNDDKQFDLKFYGFRGRDLDN
                    GHSLTFKRIQNTDEGAMIVNHNTTQAANVTITGNESIVLPNGNINKLDYRKEIAYNG
                    WFGETDKNKHNGRLNLIYKPTTEDRTLLLSGGTNLKGDIIQTGKSLKFTSGRPTPHAYN
                    HLNKRWSEMEGIIPQGEIVWDHWINRFFKAENFOIKGGSAVVSRNVSSIEGNMTVSNN
                    ANATFGVVPNQONTICTRSDWTGLTTCQKVDLTDRVINSIPKTQINGSINLTDNAIPA
                    NVKGLAKINGNVNTLINHSQFPLSNNATQIGNIRLSDNSPATVDNANLNGNHITDSAQ
                    FSLKNSHSHQIQGDKGTTVTLENAATWIMPSDTTLQNLTLNNSTILNSAYSASSNNT
                    PRRSLETETTPTSAEREHRLNITVNRKLSQGGTFQPTSSLFGYKSDKLKLSNDAEGDY
                    LSVRNTGKEPEIELEQILVESKDNQPLSDKLKFTLENDVADAGARLYKLVKNDGEFR
                    LHNPIKEQELHNDLVRAQAERLEAKQVEPFAKTQTGEPKVRSRAAARAAAFPDTLPD
                    QSKRDLAQEEAEKQKRQKDLIRSYNSALSELSAVSNSMLSVQDELDRLFVDQAQSA
                    VWINIAQDKRRIDSDAFRAYQQKTNLRQIGVQKALANGRIGAVFSHSRSDNTPEQVK
                    NHAILTMSGSFAQYQWGEDLQFGVVNVGIISAKMAEEQSRKIHRAINGVNASYQFR
                    LGQLGIQPFYGVNRYFIERENYQSEEVRVKIPSLAPNRYNAGIRVDYTFTPTDNISVK
                    PYFEVNYVDVSVNANVQTTVNLIVLQQPFGRYWQKEVGLKAEILHFQISAFISKSQGSQ
                    LGKQONVGYKLGYRW"
BASE COUNT   1496 a    776 c    890 g   1155 t
ORIGIN
```

```
Query Match        62.0%;   Score 2995;   DB 1;   Length 4317;
Best Local Similarity    82.4%;   Pred. No. 0;
Matches 3582;  Conservative   0;  Mismatches 645;  Indels 120;  Gaps   8;
```

```
Qy    253 TCAATAGTCGTTAACCACGTATTTTTTAATACGAAAAAATTACTTAATTAAATAAAACATT 312
          ||||||||||||||||| |||||||||||||||||| ||| ||| |||||||||||||||
Db      1 TCAATAGTCGTTTAACTACGTATTTTTTTAATACGAAAAAATTACTTAATTAAATAAACATT 60

Qy    313 ATGAAAAAAACTGTATTTCGTCTCGAAATTTTTAAACCGCCTTGCATTTCATTAGGATAGTA 372
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ATGAAAAAAACTGTATTCGTCTCTTAAATTTTTAAACCGCCTTGCATTTCATTAGGATAGTA 120

Qy    373 TCGCAACGCTGGGCAGGTCATCACTTATTTATTTGGGATTGACTACCAATAATTATCGTTAGTT 432
          |||||||||||||||||||||||||||||||||||| |||||| |||||||||||||||||
Db    121 TCGCAACGCTGGGCCGTGGGTCACACTTATTTATTTGGGATTGAATGATTACCAATAATATCGTGATTTT 180
```

```
Db    181 GCCGAGAATAAAGGGAGGAGTTCACAGTTGGGCGGGCTCAAAATAATTAAGGTTTATACAAACAA 240
          ||||| |||||||||||||||||||||||||||| | |||||||||| |||||||||||||||
Qy    493 GGAAATTTAGTTGGCACATCAATCAGACAAAAACCCCAATGATTGATTGATTTTTCCGGTGTGTCG 552
          ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GGGCAATTAGTTGGCCAATCAATCAATGACAAAAAGCCCGGATGAGTTGATTGGTGAGTGTGGCACAAAATGTA 300

Qy    553 CGAAATGGGGTGGGCGGGGCCATTGGTGGGGGCCATACGTATATTCATGGGGGCACAAAATGTA 612
          || |||||||| ||||||| |||||||||||||||||||||||| |||||| |||||||||||
Db    301 CGTAACGGGCGTGGGCAGCCCTTGGTTGAAAAAATCAATAATATATTGTGAGCGGTGGCACAAAACGTA 360

Qy    613 GGCTATACCAAATGTGGGATTTTTGTGCTGCGAAGGACAAAAAATCCTGACCAAACATCGTTTTACT 672
          || | |||||||| ||||||||||||||| ||||||||||||||||||||||||||||||||||
Db    361 GGAATAACAGATGTTGATTTGGGTGCAGAGGGAAACAACCCGGATCAACACACAGTTTTACT 420

Qy    673 TATAAAAATGTGAAACGGAATAATTATATAAAACGATAAAGGCCATCCTTATGAGAAAGAC 732
          |||| || ||||||||||||||||||||||||||||||||| |||| ||||||||||||||
Db    421 TATAGAGATTGTAAAACGATTACAAAAAAGATAATTACATCCTTATGCAAGACGAT 480

Qy    733 TACCACACACCACCGCCTTACATAAATTGTTACGGAGGAGACCCACCCCAATACGAATATAACTTCT 792
          |||| ||||||||| |||||||||||||||||||||||||| |||||||||||||| ||||
Db    481 TACCATAAATCCACGATTACATAAAATTCGTTACGAAGCGGCTCCAAATTGATAGACTICG 540

Qy    793 GATATGAAACGGCACACAAAATATACAGTAGGAGCGAAAATATCCCGACACGACGTCCGTATCGGC 852
          |||| ||| || | ||||||||||||||||||||||||||||||||||| ||||||||||||||
Db    541 AATATGAAATGGCAGTACTTATTCAAGATAGAACAAAATATCCAGAACGTGTTCGTATCGGC 600

Qy    853 TCCGGGTGGCAGTTTTGCGGAATAGTCACGAATACACAACGGCGACCAAAGTTGCGGGCCATAT 912
          |||| |||||||||| | |||||||||||||||||||||||||||||||||||||||||||
Db    601 TCTGGACGGCAGTTTTTGGGCGGAATAAAGCGGCCGACCAAGGTGCCGGGTCGCATAT 660

Qy    913 CATTACCTGACAGCAGGAAATACACAACCAAGGGAGGGGGCTGGTAAGT 972
          |||||||||||||||| | |||||| ||||||||||||||||| |||||
Db    661 CATTAGCTGACAGGTGGCAATACACAACAGCCGTGGAAGGAGGATAATTCGTAT 720

Qy    973 CTGACGCGGCATGTGCCGCCAAGCGGGCAAAGTTACGGCGCCCCATTCCTTATTGCAGGAGCCAAGC 1032
          |||||||||||||||||||||||||||||||||||||| | |||||||||||||||||||||||
Db    721 TTGGGAGGCGAGTCGTAAGCAGATAGGTCCATTACGGATTGCAGGGCTCAAAG 780

Qy   1033 GGCGACAGCGCGTTCGCCTACGTTGATTTATTTTTTGATGCGGAAAACAAAAATGGTTGATTTAAC 1092
          |||||||||||| ||||||||| |||| ||||| | |||||||||||||| |||||||||||||
Db    781 GGGGACAGTGGTTCTCCGATGTTTAATTGAGATAGGTGCAAAGAAAAATGGTTTAATTAAT 840

Qy   1093 GGCGTATTGAGGACGGCAACCCTTGGGCGGGGGGACACAAGAATACATCCAACTGGTGTACGC 1152
          ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    841 GGGATATTACGGGAGGCAACCCTTTGAGGCACAAAGAAAAATGGTTTCAATTGGTCGC 900

Qy   1153 AAGTCTTTTTGATGAAAACCTTGAAAAAGAATTGCTACACACGTTTTATAGCCCATCG 1212
          ||||||| |||| || |||||||||||| ||| | |||| |||||||||||||||||
Db    901 AAATCTTATTTGATGAAAATTTTGAAACAGAGAATTACATCACTTTACACCCGAGCT 960

Qy   1213 GGCAATGGTCATCACACCATCACGACAGACAAAAGCGACGCGCAGGGCCATTGTGCAAAACACA 1272
          ||||||||||||||| | ||||||| || |||| ||||||||||||||| | ||||||||||
Db    961 GGTAATGGAGTGTACACAACAAATTAGTGAAATGAATGGTCAGGGTCTTATAACTCAGAAA 1020

Qy   1273 ACAGGGAAGAACCATCGAAGTCCGGCATCCGGTTTAAAAGACGACAAATTACCCGTCCGGAAGGT 1332
          ||| || |||||| | |||||||| | ||||||||| |||||||||||||| |||||||||
Db   1021 TCAGGGATTACCATCAGAAAATAAAAATTACGTTAGCAAAATATGAGTTTACCCTTTGAAAGAG 1080

Qy   1333 AAAGACGGATGTTTNCCA------ATACCAAGGTCCAAATATATATATTACCTGCCCGTTTGAAT 1386
          ||||||| |||||||| ||      ||| |||||| | |||||||||||||||||||||||||
Db   1081 AAGGATAAAGTTCATAATCCTAGAATATGACGGACCTATATTTATTCTCCAGTTTAAC 1140

Qy   1387 AACGGTGGAAACCTGATTTCGGAGATCAAAAAAACGGCACTGTTACCTTATCGAACCAAC 1446
          ||||||||||| ||||||||||||||||||||| || ||||||||||||||| |||||||
Db   1141 AATGGAGAAAACGCTATATTTTTTTTATGGGATAAAAAACAAGTAATACCCGATGGTGCCATCTGAC 1200

Qy   1447 ATCAACCACAAGGTTGCGGCGGGCGGTTGTTGTATTTGAGGGTAACTTTACGGTATTCTCGAAAAT 1506
          ||||||||||| |||| |||||||| |||| |||||||| ||||||| |||||||||||||||
Db   1201 ATTAACCAAGGCGCGGGCGTGGGTCTTTATTTATTTGAGGGTAATTTGCAGTATCTCCAAATCT 1260

Qy   1507 TCGCAAGCGTGGGCAGGTCATCACACATTATTTATTTGGAGTGCATGGGAGTCGTTGCGAGTGAAA 1566
          || |||||||||||||| | | ||||||||||| |||| ||||||||||||||||| ||||||
Db   1261 AACCAAACTTGGCGAGGAGCTGGGCATCACGATCGTAAGTGAAAATAGCACCGTTACTGGGAA 1320
```

```
LOCUS       AR157320     4319 bp    DNA     linear   PAT 17-OCT-2001
DEFINITION  Sequence 1 from patent US 6245337.
ACCESSION   AR157320
VERSION     AR157320.1  GI:16218252
KEYWORDS
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   (bases 1 to 4319)
  AUTHORS   ...J.W. III and Falkow,S.
  TITLE     ...herence and penetration proteins
  JOURNAL   Patent: US 6245337-A 1 12-JUN-2001;
FEATURES             Location/Qualifiers
     source          1..4319
                     /organism="unknown"
BASE COUNT     1497 a    776 c    891 g   1155 t
ORIGIN
```

RESULT 2

Query Match   61.5%;   Score 2970;   DB 6;   Length 4319;
Best Local Similarity 82.3%;   Pred. No. 0;
Matches 3581; Conservative 0; Mismatches 645; Indels 124; Gaps 10;

```
Qy    253 TCAATAGTCGTTAACCACGTATTTTTAATACGAAAAATTTAATAAACATT  312
          ||||||||||||| ||||  | |||||| |||| || || |||||||||
Db      1 TCAATAGTCGTTAACTA-GTATTTTTTAATCGAAAAATTACTTAATAAACATT  59

Qy    313 ATGAAAAAAACTGTATTTTCGTCTGAAATTTTTTTTAACCCGCTTGCAATT  372
          |||||||||||   ||| |  ||| ||||| ||||||   || |||||||
Db     60 ATGAAAAAAACTGTATTTTCGTCTTAAATTTTTTTAACCGCTTGCATTAGTA  119

Qy    373 TCGCAAGCGTGGGCAGGTCATACTTATTTTGGGATTGACTACCAATAATAT  432
          |||||||||||||||  ||| ||||||||||||||| ||||| |||||||
Db    120 TCGCAAGCGTGGGCAGGTCATACTTATTTTGGGATTGGACTACCAATAAT  179

Qy    433 GCCGACAATAAAAGGAAGTTCACAGTTGGGGCGTAAAAA  492
          ||||||||||||||||||||||||||||||||||||||
Db    180 GCCGAGAATAAAAGGAAGTTCACAGTTGGGGCGTAAAAA  239

Qy    493 GGAAATTAGTTGGGCACATCAATGACAAAAAGCCCCAA  552
          ||||||||||| ||||||||||| |||||||||||||
Db    240 GGGCAATTAGTTGGGCACATCAATGACAAAAAGCCCAA  299

Qy    553 CGAAAAAGGGGTCGGCGGGGGCATTGGGGAGTGAAAGTG  612
          |||||||| ||||||||||||||||||||| ||||||
Db    300 CGTAAGCGGTGGGCAGGCCTTGGTTGAAATCAATAT  359

Qy    613 GGCTATTACCAATGTGGATTTTTGGTTCTGAGGACAAAG  672
          |||||||||||||||||| ||||||||||||||||||
Db    360 GGAATTACAGATTGAATTTTGGGTCGAGATAATTTTT  419

Qy    673 TATAAATTGTGAAACGGAATAATTTATAAAAAACGGAATA  732
          |||||||||||| ||| ||||||||||| |||||||||
Db    420 TATAAGATTGTAAAACGAAATAACTACAAAAAA  479

Qy    733 TACCACAACCCACGCTTACAATAAATTTGTTACGGAAGCCA  792
          ||||||||||||||||||||||||||||| |||||
Db    480 TACCATAATCCAACGATTACAATAAATTCGTTACA  539

Qy    793 GATAAGGAGGCGATGCCCCAAGCGGCCAATTACGGCCCCAT  852
          ||||||||||||||||||||||||||||||||||||
Db    540 AATGAATGGCAGTACTTTATTTATTTATGATGGCGA  599

Qy    853 TCCGGGTGGCACTTGGCGAAACGGATCAAAAACCGGACCAA  912
          ||||||||||||||||||||||||||||||||||||
Db    600 TCTGGACGGCAGTTTGGCGAAATGATCAAGACAAA  659

Qy    913 CATTACCTGACAGCAGGCAATACAACCACCAAGGCGGAG  972
          |||||||||||||||||||||||||||||||||||
Db    660 CATTAATCTGACAGCGAATACAATCACGCGTGGAGC  719

Qy    973 CTGAGCGGCGATGTGCCCCAAGCGGCGGCCAATTACGG  1032
          |||||||||||||||||||||||||||||||||||
Db    720 TTGGGAGGCGATGTTCGTAAAGCGGGAGAATATGGT  779

Qy   1033 GGCGGACAGCGGTTCGCCCTATGTTTATTTATGATGCGG  1092
          |||||||||||||||||||||||||||||||||||
Db    780 GGGGACAGTGGTTCTCCGATGTTTATTTATGATGCT  839

Qy   1093 GGGCTATTGAGGACCGCAACCCTTGGGCCGGGACAGAGA  1152
          |||||||||||||||||||||||||||||||||||
Db    840 GGGATAATTACGGGAAGGCAACCCTTTGAGGCAAGA  899

Qy   1153 AAGTCTTTTTTGATGAAATCCTTGAAAGGAAATTCC  ...
```

```
Db   4234 CGTTGGTGTAAAAATCAACATAATTTTAT  4260

Qy   4540 CGTTGGTGTAAAAATCAACATAATTTTTAT  4566
          ||||||||||||||||||||||||||||||
Db   4174 TCAAAATCTCAAGGTTCACAACTCGGCAAACACGAAAAAGTGGGCGTGAAATGGGCTAT  4233

Qy   4480 TCTAAATCTCAAGGTTCGCAACTCGGTATCAGGCAGAATATGGGGCGTGAAATTAGGATAT  4539
          |||| ||||||||| |||| |||| ||    |  |  ||||| |||||||||| ||||||

Db   4114 TGGCAAAAAGAAGTGGGATTAAAGCAGAAATTTAACATTCCAAATTCCGGCTTTATC  4173
Qy   4420 TGGCAAAAAGAAGTGGGATTGGGAGAAAATTTACAATTCCAACTTTCTGCTTTTTATT  4479
          |||||||||||||||||||| ||| | |||||| |||||||| |||| |||||| ||

Db   4054 AACGCTAACGTACAAAACCACGGTTAAATCTCACCGGTGTTGCAACAACCATTGGACGTTAT  4113
Qy   4360 AACGCTAACGTACAAAGCCACTCGTAAAATCGGCGCGGTGTTGCAACAACCATTGGACGTTAT  4419
          |||||||||||||| |||| ||| |||||| |||||||||||||||||||||||||||||

Db   3994 ACTCCGACAGATAAATATCAGCGGTTAAGCCTTATTTCGTCAATAGTTGATTATACAGAT  4053
Qy   4300 ACCCCGACAGGAAATCACGCCGTTAAGCCTTATTGTGTGATTAGTTGTGATATACATTTCA  4359
          ||  |||||| ||||| ||||||||||||||| | | ||||| |||||| |||||| |

Db   3934 AAAACGCCTACCTTGCATTAAATGCTATAAAGCTGGAGTTGATTATATACATTT  3993
Qy   4240 CAAACACGAGGCCTTGCATTTAATGCTTATAAATCTGGAGTACGGGTCGGATTACAGTTGG  4299
          ||||| |||| |||||||||| |||||| ||||| |||||||| ||  |||| || |||

Db   3874 TTTGGAGTTAATCGCTATTTTTAATTTGAAACGTGAAAATATCAAGAGGAAGTGAGAGTG  3933
Qy   4180 TTTGGAGTTAATCGCTATTTTTTATTTGAAAATATCAATCTGGGAGAAAGTG  4239
          |||||||||||||||||||| ||| |||||| ||||| |||||| ||||||

Db   3814 AATTAAGGCCTGAATGCAAGTTAACAGTTCCGTTTAGGGCAATTGGGCCATCAGCCTTAT  3873
Qy   4120 AATTATGGCGTGAATGCAAGATTAATCGTTCCATTAGGCGTATTCGATTCAGCGCTTAT  4179
          ||||| |||| ||||||||| || || ||||| || ||| ||| ||||||| | ||||

Db   3754 GGAATCAGTCGCGAGTGTAAAAATGGCTGAAGACAAAAAGCCGAAAAATTACAATTGGGCG  3813
Qy   4060 AATTATGGCGTGAATGGTGTGAAGAACAAAAGCCGAAAAATTCATCGAAAAACGATA  4119
          |||| ||||| ||||| | ||||| |||||| || |||||||  || |||||  |||

Db   3694 ATGTCGGGTTTTGCCCAATATCAATGGGGTGTAAACGTGGGAACG  3753
Qy   4000 ATGTCGGGTTTTGCCCAATATCAATGGGTGATTACAAATTTGGTGTAAACGTGGGAACG  4059
          ||||||||||||||||||||||||||| |||  |||| ||| |||||||||||||||

Db   3634 CATAGCCGTTCAGATAATACCTTTGATGAACAGGTTAAAAATCACGCGACATTAACGATG  3693
Qy   3940 CATAGCCGTTCAGATAATACTTTTGATGAACAGGTTAAATTCACCAACATTAACGATG  3999
          |||||||||||||||||||| |||||||||||||||||  ||| |||||||||||||

Db   3574 TTACGTCAAAATTGGGCGTGCAAAAAAGCCCTTAGCTAATAGGGACGAATTGGGGCAGTTTTCTCG  3633
Qy   3880 TTGCGTCAAATTGGGGGTGCAAAAAAAGCCCTTAGCTAACGGACGAATTGGGGCAGTTTTCTCG  3939
          || |||||||  |||| ||||||||| |||||||||||||  |||||||||||||||||||||

Db   3514 GCACAGGATAAAAACGCTATGATTCGGTTCCGTGCTTATCAGCGAGAAAACGAAC  3573
Qy   3820 TCACAGGATAAAAGACGTTATGATTCGGTTCCGTTGCTTATCAGCAGAAAACGAAC  3879
          | |||||||||| | ||||||||||||||||||| ||||||||| ||||||||||

Db   3454 CAAGATGAAGTAGATCGTGCTTTTTTGTGATCAAGCAATCTGCCGTGGGTGACAAATATC  3513
Qy   3760 CAAGATGAAGTAGATGCGTCTTTTTTGTAGATCAAGCTCAAATCTGCGGTGTGGACAAAATATC  3819
          |||||||||||||||  | ||||||||| |||||||| | ||||||  |||| |||||| ||||

Db   3394 CGTTATTCAAATAGTAGTGCGTTTATCGAGAATTATCTGCAACAGTAAAATAGTAATGCTTTCTGTT  3453
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy  3343 ---AGAGCGGGCCGTTTCTTGATACCCCGCGATCAAAGCCCAGTTATAAAGCGCAATACAAGCC 3399
              |||| ||| ||||||||| |||   || |||   |||  ||| ||| |||||| || ||||
Db  3117 GGGAGAGCGAGCGTTTCTTCCTGATCATCAAAGCCCTGCTGATCGATCAAAGCCAATAGAAGCC 3176

Qy  3400 GAACTCGAGACGAGCGTTTCCTGATCACCCTGCCTGATCAAAGCTCGAGAAAGTA 3459
Db  3177 AAACAAGCTGAACTGACTGCTGAAACACAA------AAAGTAAGGCAAAAACAAAAAG 3231

Qy  3460 ACTGCACTTAACCAAAAGAACGAGCGAAGTAAAAACACTCAAGATAAAGCAAATTAGTC 3519
Db  3232 TCCGGTCAAAAAGAGCAGTGTTTTC----TGATCCCTGCTTGATCAAAGCCTGTTCGCA 3287

Qy  3520 TTGGCAACTGCATTGGTGGAAAAAGAAACCGCTCAGATTGATTTTGCTAATGCAAATTA 3579
Db  3288 TTAGAAGCCGCACTTGAGGTTATTGATGATGCCCAC------------------------- 3321

Qy  3580 GCTCAGTTGAATTTAACACAACAACTAGAAAAAAGCCTTAGCAGTGGCTGCAAGCAGGA 3639
Db  3322 -------------------------------------------------AGCAATCGGAA 3332

Qy  3640 AAAGAGACGCTAAAGCTCAAGAGACAAGCACGCAAAAAAGACAAAAAGACTTGATCAGC 3699
Db  3333 AAAGAATCGTCTACGCTTCAAGAGAAAAAGACGCGAAAAAACAACAAAAAAGACTTCAGC 3392

Qy  3700 CGTTATTCAAATAGTGCGTTATCAGAATTATCTGCAACAGTAAATAGTAGTATGCTTTCCGTT 3759
Db  3393 CGTTATTCAAATAGTGCGTTATCGCAACGAGTAAATAGTTGGCTTCGTGTT 3452

Qy  3760 CAAGATGAATTAGATCGTCGTCTTTTTTTTGTAGATCAAGCTCAATCTGCCGGTGTGGACAAAT 3819
Db  3453 CAAGATGAATTAGATCGTCGTCTTTTTTGTAGATCAAGCACAATCTGCCGGTGTGGACAAAATC 3512

Qy  3820 TCACAGGATAAAAAGACGTTATGATTCGATGCGTTCCGTCGTTCCGTGCCTTAT---CAGCAGAAAACG 3876
Db  3513 GCACAGGATAAAAAAGACGCTAATGAATCTGATCGCGTTCCCGTGCTTATCGCAGCAGAAAAAACG 3572

Qy  3877 AACTTGCGTCAAAATTGGGGTGGGGTGCAAAAAAACCTTAGCTAACGGACGCAAGTTGGGGCCAGTTTTC 3936
Db  3573 AACTTACGTCAAAATTGGGGTGCAAAAAAGCCTTAGCTAACGGGAATTGGGGCCAGTTTTC 3632

Qy  3937 TCGCAATGGCGGTTCAGATAATACTTTTGATGACAGGTTAAAAATCACGCAACCATTAACG 3996
Db  3633 TCGGCAATAGCCGTTCAGATAATAACCTTTGATGACAGGTTAAAATCACGCGACCACATTAACG 3692

Qy  3997 ATGAAGTGTCGGGTTTTTGCCCAAATATCAAATGGGGGTGGATTTTACAAATTGGGTGAAAACGTGGGA 4056
Db  3693 ATGAAGTGTCGGGTTTTTGCCCAAATATCAAATGGGGGTGGATTTTACAAATTGGGTGTAAACGTGGGA 3752

Qy  4057 ACGGGAATTAGTGCGGTAAAATGGCTGAAGACAAAGCCGAAAATTCATCGAAAAAGCG 4116
Db  3753 ACGGGAATCAGTGCGGTAAAATGGCTGAAGACAACCGAAAAATTCATCGAAAAACG 3812

Qy  4117 ATAAATTATGCGTGAAATGCAAGTGTTATTCGTTCCATTTAGGCAATTGGGTATTCAGCCCT 4176
Db  3813 ATAAATTATGGCGTGAAATGCAAGTTATCCGTTCGGTTTAGGCAATTGGGCATTCAGCCAA 3872

Qy  4177 TATTTTGGAGTTAATCGCTATTTTTATTGGAACGTAAAAATATATCAAATCTGAGGAAGTGAAA 4236
Db  3873 TATTTTGGAGTTAATCGCTATTTTTATTGAACGTGAAAAATTATCAATCTGAGGAAGTGAGA 3932

Qy  4237 GTGCAAACACCGAGCCTTGCATTTAATCGGTATTAATGCTGGGTACGGGTCGATTATACG 4296
Db  3933 GTGAAAAACGCCTAGCCTTGCATTTAATCGGTATTAATGCTGGGTACGGGTCGATTATACA 3992

Qy  4297 TTTACCCGCACGAGAGAATATCAGCGGTTAAGCGTTAATCGGCGGTTGCAACAACCATTGGACGT 4356
Db  3993 TTTACCCGCAGAGAATAATATCAGCGGTTAAGCCCTTATTCTTCCGTCAAATTAGCGGTGTGATGTT 4052

Qy  4357 TCAAACGCTAACGTCACAAACCACTGTAAATCGCGCGGTGTTTCGCGGGTTGCAACAACCATTTGGGACTT 4416
Db  4857 GTGGGATTAAAGCGGGAAATTTTACATTCAACTTCTGCCTTTT 4476

Db  4113 TATTGGCAAAAAGAAGTGGGATTAAAGGCAGAAAGTTTACATTCCAAATTCCGCTTTT 4172
Qy  4477 ATTTCTAAATCTCAAGGTTCCGCAACTCGGGTAAACACGCAGAAATATGGGCGTGAAATTGGGC 4536
Db  4173 ATCTCAAAATCTCAAGGTTCACAACTCGGCAAACAGCGCAAAATGTGGGGCGTGAAATTGGGC 4232

Qy  4537 TATCGGTTGGGTAAAAATCAACAAATTTTAT 4566
Db  4233 TATCGGTTGGGTAAAAATCAACATAATTTTAT 4262
```

RESULT 3
U32710/c

```
LOCUS       U32710               13818 bp    DNA     linear   BCT 29-MAY-1998
DEFINITION  Haemophilus influenzae Rd section 25 of 163 of the complete genome.
ACCESSION   U32710 L42023
VERSION     U32710.1  GI:1573200
KEYWORDS    .
SOURCE      Haemophilus influenzae Rd.
  ORGANISM  Haemophilus influenzae Rd
            Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
            Haemophilus.
REFERENCE   1  (bases 1 to 13818)
  AUTHORS   Fleischmann,R.D., Adams,M.D., White,O., Clayton,R.A.,
            Kirkness,E.F., Kerlavage,A.R., Bult,C.J., Tomb,J., Dougherty,B.A.,
            Merrick,J.M., McKenney,K., Sutton,G.G., FitzHugh,W., Fields,C.A.,
            Gocayne,J.D., Scott,J.D., Shirley,R., Liu,L.I., Glodek,A.,
            Kelley,J.M., Weidman,J.F., Phillips,C.A., Spriggs,T., Hedblom,E.,
            Cotton,M.D., Utterback,T., Hanna,M.C., Nguyen,D.T., Saudek,D.M.,
            Brandon,R.C., Fine,L.D., Fritchman,J.L., Fuhrmann,J.L.,
            Geoghagen,N.S., Gnehm,C.L., McDonald,L.A., Small,K.V., Fraser,C.M.,
            Smith,H.O. and Venter,J.C.
  TITLE     Whole-genome random sequencing and assembly of Haemophilus
            influenzae Rd
  JOURNAL   Science 269 (5223), 496-512 (1995)
  MEDLINE   95350630
   PUBMED   7542800
REFERENCE   2  (bases 1 to 13818)
  AUTHORS   Tatusov,R.L., Mushegian,A.R., Bork,P., Brown,N.P., Hayes,W.S.,
            Borodovsky,M., Rudd,K.E. and Koonin,E.V.
  TITLE     Metabolism and evolution of Haemophilus influenzae deduced from a
            whole-genome comparison with Escherichia coli
  JOURNAL   Curr. Biol. 6 (3), 279-291 (1996)
  MEDLINE   96398784
   PUBMED   8805245
REFERENCE   3  (bases 1 to 13818)
  AUTHORS   White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUL-1995) The Institute for Genomic Research,
            Medical Center Dr, Rockville, MD 20850, USA
REFERENCE   4  (bases 1 to 13818)
  AUTHORS   White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-SEP-1997) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
  REMARK    The H. influenzae sequence has been updated by R. Fleischmann. New
            database matches have been assigned, product names have been
            improved, and a number of frame shifts have been corrected. We
            gratefully acknowledge the work of Tatusov et. al. We have
            incorporated their annotation into the /notes fields of the
            corresponding H. influenzae genes
REFERENCE   5  (bases 1 to 13818)
  AUTHORS   White,O., Clayton,R.A., Kerlavage,A.R., Fleischmann,R.D.,
            Peterson,J., Hickey,E., Dodson,R. and Gwinn,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAY-1998) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
  REMARK    The whole genome was shifted by 588 nucleotides for a new start
            On Sep 30, 1996 this sequence version replaced gi:1220911.
COMMENT
FEATURES             Location/Qualifiers
     source          1..13818
                     /organism="Haemophilus influenzae Rd"
```

Qy   125 AAGTATTAAGCAAAAACCTAGAAATTTGGCTTAATTACTATATATAGTTCATTTAT 184
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 13818 AAGTATTAAGCAAAAACCTAGAAATTTGGCTTAATTACTATATATAGTTTACTCATTTAT 13759

Qy   185 TTTCTTTTGTTGCCTTTAGTTCGTTTTTTTTTAGGTGAAATCACCGGACT 244
          |||||||||||||||||||||||||||||||||| ||||||||||||||||
Db 13758 TTTCTTTTGTTGCCTTTAGTTCG-TTTTTTAGGTGAAATCCCTTAGAAAATCACCGACT 13700

Qy   245 TTTATTGTTCAATAGTTCAATTAGTTCAATATATTTTTTATTTAATTAAA 304
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 13699 TTTATTGTTCAATAGTCGTTTTAACCACGTATTTTTTTAATTACGAAAAATTACTTAATTAAA 13640

Qy   305 TAAACATTATGAAAAAAAACTGTATTTTGTTCTCTGAATTTTTTTTCATTTCATTAG 364
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 13639 TAAACATTATGAAAAAAAACTGTATTTTTCGT-TTTTTTAAGCTGAAATCACCGACT 13580

Qy   365 GGATAGTATCGCAAGCGTGGGCAGGTCATCATTTATTTTGGGATTGACTACCAAATATATC 424
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 13579 GGATAGCATCACAAGCCCTGGGCAGGTCAACTTATTTTGTCATTTTTTTCATTTATC 13520

Qy   425 GTGATTTTGCCGGAATAAAGGGAATTCACAGTTGACTACCAAATATATTTACA 484
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 13519 GTGATTTTGCCGGAATAAAGGGAAGTTCACAGTTGGGGCTAAAAAATATTGAGGTTTACA 13460

Qy   485 ATAAAAATGGAAATTTAGTTGGGCACATCAATGACAAAAGCCCCAATGAGTTTTTTCCG 544
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 13459 ACAAAAGAAGGGCAATTAGTTGGCACAATGACAAAAGCCCGATGAATTGATTTTTCCG 13400

Qy   545 TGGTGTCGCGAAATGGGGTGGGGCCATTGGGGTGGGGCAC 604
          |||||||||||||||||||||||||||||||||||||||
Db 13399 TGGTGTCGCGTAACGGGGTGGGGCCATTGGTAGGGGACATTTTGCTGAGGACATTTGGAC 13340

Qy   605 ATAATGTAGGCTATACCAAATGTGAATTTTTGTGTCTGAAGGACAAAATCCTGATCACATC 664
          |||||||||||||||||||||||||||||||||||||||||||
Db 13339 ATAACGGGGGATATAACCATGTTGAATTTGGGTCGAGAAGGAGAAAACCCTGATCAGCACC 13280

Qy   665 GTTTTACTTTATAAAATTGTGAAACGGAATAATTATATAAA---AACGATCAAACGCATCCTT 721
          |||||||| ||||||||||||||||||||||||||||||| ||||||||||||||||||
Db 13279 GCTTTACTTTACAAAATTGTAAAAGAAATTAATCAGCTTGGGAGAAAGCATCCTT 13220

Qy   722 ATGAGAAAGACTACCACAACCCACGCTTACATAAAATTTGTTACGGAAGCCACCCCAATCG 781
          |||||||||||||||||||||||||||||||||||||||||||||||
Db 13219 ATGATGGAGATTAATCATAATGCCTCGTTTTACATAAAATTTTATTCGGCTTCCGTG 13160

Qy   782 ATATGACTTCTGATATGAACGGCAACAAAATAATACAGATAGGACGAAAATATCCCGAAGCG 841
          |||||||||||||||||||||||||||||||||||||||||||||||
Db 13159 GTAAGACAACAAATATGATGGAAAAGTATATGCTGATAGAGAAACTATCCTGAGGCGTG 13100

Qy   842 TGGCTATCGGCTCCGGGGTGGGCAGTTTGGGGAAACGATCAAAAACAAACG------ 889
          |||||||||||||||||||||||||||||||||||||||||||
Db 13099 TACGTATAGGCTCAGGAGCGTCAGTATTGGCGTACAGATAAAGATGAAGCAAACGAATGTAC 13040

Qy   890 ------GGACCAAGTTGCCGGGGCCATAATCATTACCTCGACAGCAGGCAATACACAACC 943
                |||||||||||||||||||||||||||||||||||||||||||||
Db 13039 ATAGTTCATATTATGTCCAGGTGCATAATCGTTATCTTACTCGAGGAAAATACCCATACTC 12980

Qy   944 AAGGCGGGAGCAGGGGGCCGGCGGCTGGTGTCAAGTCTCGAAGTGCGGCCAAGCGGGCAATT 1003
          |||||||||||||||||||||||||||||||||||||||||||||||
Db 12979 AGATGGGAAAATGGTAAATGGTACACTCAATCTTAGTGGTTAGTTAGCCCTAATCATT 12920

Qy   1004 ACGGCCCCCATTCCTATTGCCAAGGCTCAAGCGGCCGACAGCGGTTGCGCCTATTGTTATTTATG 1063
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 12919 ATGGTCCATTACCAAACGGGTGGTTTCTAAAGGGGATAACCGGTTCGCCCAATGTTTATTTATG 12860

Qy   1064 ATGGGGAAAAACAAAAATGGTTGATTAACGGCGTATTGAGGACCGGACCGGGGG 1123
          ||||||||||||||||||||||||||||||||||||||||||||||
Db 12859 ATGGGGAAGAAGAAACAATGGCTTAATAAATGGCTTTAATTACGTGAAGAATGGGTTTCG 12800

Qy   1124 GGACAGAGAAATACAATTCCAAACTGGTACG---CAAGTCTTTTTTTGAGAAATCCTTGAAA 1180
          ||| || || || || ||||||||||||| |||||||||||||
Db 12799 GAAGGGTAATGGGTTCAGTTAATACGTGAAGAATGGGTTTTATAATACGTGAAGTTCTTGCGG 12740

Db 12739 TTGATACCCCTAGTGTTGTTTTCAACGCTATAATTCCCCAATAAATGGACATTATTCCTTTG 12680
          ||| ||||||||||||||||||||||||||||||||||||
Qy   1235 CAGACAAAGGCGACGGCGAGCGGCATTGTCA----------------AACAACAAA 1273

Db 12679 TATCAAATAAATGGATGGCTAACAGGTAAATTAACTTAACTAGAGACTAGTAGAGATGGGCTCTA 12620
          || |||||| ||||||||||||||||||||||||||
Qy   1274 CAGGAAGACCATCTGAAGTCCCATCGGTTTAAAAGAGACGAAATTACCTGCCGAAGGTA 1333

Db 12619 AAGCAAAATCAGAAGTAGGAACTGTGAAGTTATTTAATCCATCGTTAAATCAAAACGCTA 12560
          ||| ||||||||||||||||||||||||||||||||||||
Qy   1334 AAGACGATAGTGTTTACCAATACCAAGGTCCAAAAATATACCTGCCGTTGTTTAAATAACGGTG 1393

Db 12559 AAGAAACATGTTAAAGCAGCAGGCAGGCTATAATATTTACCAACCAAGAATGGAATARGGAA 12500
          ||||||||||||||||||||||||||||||||||||
Qy   1394 GAAACCGTATTCGGAGATCAAAAAAACGGCACTGTTACCTTATCAACCAACATCAACC 1453

Db 12499 AAAATATTTACCTTGCCGACCAAGGAAAAAGGAAACTTTAACAAATCAAAATAATATCAAATC 12440
          |||| |||||||||| ||||||||||
Qy   1454 AAGGTGCGGCGGCGGGTTTGTTATTTTGAGGGGAAGGAAGTTTAACAAATAAATTAAAG 1510

Db 12439 AAGGTGCGTGGTGGGGATTATACTTTGAAGGTAATTTGTTGTAAAAGCCAAGCAAAATAATA 12380
          ||||||||||||||||||||||||||||||||||||
Qy   1511 CAACTTGGCAAGGTGCTGGGAGTGGCATGTAGGTGGAAGCAGCAGTACTGTTGTGAAAGTAA 1570

Db 12379 TAACTTGGCAAGGTGCAGGCGTATCTATTGGACAAGATGCAACTGTTGAATTGCAATGGTTC 12320
          |||||||||||||||||||||||||||||||||||
Qy   1571 ATGGTGTTGAAAATGATCGCCTTTCTAAAAATGGGCAAAAGCGCACGTTAAAGCCA 1630

Db 12319 ACAATCCTGAAAAATGATCGTGGTTTATCTAAAAATGGTATAGGCACTTTAATGGTTA 12260
          |||||||||||||||||||||||||||||||||||
Qy   1631 AAGGGGAAAAATAAAAGGTTCGGACGCGGGGTAGGCGCGTAGGTGTAAAGTCATTTTGGGAGG 1690

Db 12259 AGGGAAAGAAATTAGGAGAGTTTAAGTCGGGTAACGGCAAAGTCATTCTAGATCAACAAG 12200
          |||||||||||||||||||||||||||||||||||
Qy   1691 CAGACGATCAAGCAACAAACAAGCCTTAGTGAAATTGGCTTGGTTAGTGGCAGAGGTA 1750

Db 12199 CAGATGAAGCGGGTCAAAAAACAAGCTTCAAGATGGCTGTCGCAGCAA 12140
          |||||||||||||||||||||||||||||||||||
Qy   1751 CGGTTCAGTTAAACGATGACAAGCAATTTAATACTGTAATACCATTAAGTAGTCGTGGTTTCCGTG 1810

Db 12139 CACTTCAATTAAATAGTACAGATCAAGTTGATCCTTAACAACAATATCTATTGTTCTTGTG 12080
          |||||||||||||||||||||||||||||||||||
Qy   1811 GTGGTCGCTTAGATCTTAATGGCCATTCATTAAGTTTAAACGTATCCAAAAATACGGATG 1870

Db 12079 GTGGTCGCTTAGATCTTAACGGCCATTCATTAAGCTTAAACGTATCCAAAAATACGGACG 12020
          |||||||||||||||||||||||||||||||||||
Qy   1871 AGGGAGCAACGATTGTTAATCACAATGCCACAACAGAATCTCAGTGACCAATACTTGCA 1930

Db 12019 AGGGCGGGATGATTGTGAACCAATAACAACTCGGTAATAATTACTATTACTGGGA 11960
          ||||||||||||||||||||||||||||||||||||
Qy   1931 GGGATACCATTAATG-----------------ACAACACTGGGGATTTAACCAATAAACGTG 1975

Db 11959 ACCAAAGTATTACTGCTCCATCTAATAAAAAGAATATTAATAAAACTTGATTACAGCAAAG 11900
          |||||||||||||||||||||||||||||||||||
Qy   1976 ATATTGCTTTTAATGGTTGGTTTGGTTGGGTGATAAAAGATGATACTAAAAAATAACTGGACGTTTGA 2035

Db 11899 AAATTGCCTACAACGGTTGGGTTTGGGTTGGGCAAAACAGATAAAAAACACACAAGGACGATTAA 11840
          |||||||||||||||||||||||||||||||||||
Qy   2036 AIGTTACTTACAATCCGGCTTAACAAGATAATCACTTCCTTCTTATCAGGTGGAACAAATT 2095

Db 11839 ACTTTATTTATAAAACCAACCACAGAACGAATCGTAGCTTTTGTTGCTACTTTCAGGTTGCACAAACT 11780
          |||||||||||||||||||||||||||||||||||
Qy   2096 TAAAAGGCAAATTAACTCCAAGACGGTGGCACTTTAGTGTTTAGTGTCGCCCAAAACCAC 2155

Db 11779 TAAAAGGCGGATAATTAACCAAACAAAAAGGTAAACTATTTTTCAGGGTTAGACCGACACCCC 11720
          |||||||||||||||||||||||||||||||||||
Qy   2156 ACGCATACAATCACATTTTAAA-----TCGCCTAAACGAGCTGGGCGACCTAAGGGGCAAG 2209

Db 11719 AGGCCTACAAATCAATTTAGACAAACGGTTGGGTCAGAAATGGAGGAGGGGTATCCCACAAGGGGAAA 11660
          |||||||||||||||||||||||||||||||||||
Qy   2210 TGGTTATTGATGACGAIGGATTGGCAGACCGGTACAATTTAAAGCTGAAAACTTCCAAAATTAAG 2269

RESULT 4
LOCUS   AF517153
AF517153

LOCUS   AF517153   4054 bp   DNA   linear   BCT 08-JUL-2002
DEFINITION Haemophilus influenzae biotype aegyptius strain HK871 hap gene,
            complete sequence.
ACCESSION   AF517153
VERSION   AF517153.1 GI:21686513
KEYWORDS
SOURCE   Haemophilus influenzae biotype aegyptius.
ORGANISM   Haemophilus influenzae biotype aegyptius
            Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
            Haemophilus.
REFERENCE   1 (bases 1 to 4054)
AUTHORS   Kilian,M., Poulsen,K. and Lomholt,H.
TITLE   Evolution of the paralogous hap and iga genes in Haemophilus
            influenzae: evidence for a conserved hap pseudogene associated with
            microcolony formation in the recently diverged Haemophilus
            aegyptius and H. influenzae aegyptius
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 4054)
AUTHORS   Kilian,M., Poulsen,K. and Lomholt,H.
TITLE   Direct Submission
JOURNAL   Submitted (03-JUN-2002) Medical Microbiology and Immunology,
            University of Aarhus, The Bartholin Building, Aarhus DK-8000,
            Denmark
FEATURES   Location/Qualifiers
     source   1..4054
                     /organism="Haemophilus Influenzae biotype aegyptius"
                     /strain="HK871"
                     /db_xref="taxon:725"
                     /clone="BPF1"
                     /note="biotype: aegyptius"
     gene   1..4054
                     /gene="hap"
                     /pseudo

COUNT   1409 a   700 c   827 g   1118 t

Matches 3219; Conservative 0; Mismatches 805; Indels 292; Gaps 14;

47.7%; Score 2304; DB 1; Length 4054;
74.6%; Pred. No. 0;

Qy 4430 AAGTGGGATTAAAAGCGGGAAATTTTACATTTCTGCTTTTATTTCTAAATCTC 4489
        |||||||||||| ||| ||||||||||||| |||||||||||||||||||||||
Db 9520 AAGTGGGATTAAAGCCAGAAAATTTTACATTTCCAACTTCTGCTTTTATTTCTAAATCTC 9461

Qy 4490 AAGGTTCGCAACTCGGTAAACAGCGAAATATGGGCGCTGAAATTAGGATACGGTTGGTAAA 4549
        |||||||||||||||| ||||||| ||| |||||| ||||||||||| | |||||||||||
Db 9460 AAGGTTCGCAACTCGGCAAACAGCAAAAGTGGGGGTGCTGAAATTGGGTATCGGTTGGTAAA 9401

Qy 4550 AATCAACATAA------------------TTTTTAT 4566
        ||||||||||| |||||||
Db 9400 AATCAACATAATTGTATCGTTTATTCGATAAACAAGGTGGGGCAGATCCCACCTTTTTTAT 9341

Qy 4567 TCTAAATAATGGAACTTTATTTAATTAAAAGTATCTAAGTAGCACCC--TATAGGGGATTA 4624
        | |||||||||||||||||||| ||||||| || ||||||||||||| |||||||||||||
Db 9340 TTCAATAATGGAACTTTATTTATTTAATTAAGACCATCTAAGTAGCACCCCATATAGGGGATTA 9281

Qy 4625 ATTAAGAGGATTTAATAATTAATGAATTTAAACTAAAATTTTACCCGCATTTGCTGCAGTCG 4684
        |||||||||||||||||| ||||| ||||||| |||| ||| ||||||||||||||||
Db 9280 ATTAAGAGGATTTAATAATTGAATTTAAACTTTTACCAGCAATTTGCTGCAGTCG 9221

Qy 4685 TATTATCTCGTTGTTGCAAAAGGAATGCACCTGAAATGACAAAAATCATCTGCCCAAATAGCTG 4744
        ||||||| |||||||| |||||||| ||||||||||||||||||||||||||||||||||||||
Db 9220 TATTATCTGCTTGTTGCGAAAGGATGGCACCTGAAATGACAAAAATCATCTGCCCAAATAGCTG 9161

Qy 4745 AAAATGCAAAACACTTCCAACAATCACTGATAAAAACAGTTGTTATTCTTGCAATAAACAAA 4804
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 9160 AAAATGCAAAACACTTCCAACAATCACTGATAAAAACAGTTGTTATTCTTGCAATAAACAAA 9101

Qy 4805 CTGTGACTGCAGTGTGTATCAATTG 4828
        | |||||||| |||||||||||||||
Db 9100 CGGTGACTGCTGTGTGTATCAATTG 9077

Qy 271 CGTATTTTTAATACGGAAAAATTACTTAATTAATTAAACATTATG--AAAAAACTGTAT 328
       ||||||||||||| |||||||||||||||||||||||||||| | |||||||||||||
Db   1 CGTATTTTTAATACGGAAAAATTACTTAATTAATTAAACATTATGGAAATTTTATGAAAAAAACTGTAT 60

Qy 329 TTCGTCTGAATTTTTTAACCGCTTGCATTTCATTTCATTGGAATAGTATCGCAAAGCTGTGGGCAG 388
       |||| ||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db 61 TTTGTCTGAATTTTTTAACCGCTTGCATTTCATTTCATTGGAATAGTATCGCAAAGCTGTGGGACAG 120

Qy 389 GTCATACTTATTTTGGGATTGACTAGCCAATAATTATCGTCAATTTTTGCCGAGAATAAAGGA 448
       || |||||||||||||||||||||| |||||||| ||||| |||||||||||||||||||||||
Db 121 GTCACACTTATTTTGAGATTGACTACCAATATATGTCGTGATTTTTGCTCAGAAATAAAGGGA 180

Qy 449 AGTTCACAGTTGGGGCTAAAAATAATATTGAGGTTTTAACGATAAAAAATTTAGTTGGCA 508
       |||||||||||||||||||||||| |||||||||||||||| |||||||||||||||||||
Db 181 AGTTCACAGTTGGGGCTAAAAATATTGAGGTTTTAACAAAAGAGGGGCATTTAGTTGGCA 240

Qy 509 CATCAATGACAAAAAGCCCCAAATGATTGAGTTGATTGATTTTCGGGTCGGGAAATGGGGTGGGGGCG 568
       |||||||| ||||||| ||||| || |||||||||||| || |||||||||| |||||||||||||||
Db 241 CATCAATGACAAAAGCCACCTATGATCAGCAATTCCTGATCAGCAACCATCAAATCCCTATCAAATTGTAAAA 300

Qy 569 CATTGGTGGGGGATCAGTATATTGTGGAGTGTGGCACATAATGTAGGGCTATAACCAATGTGG 628
       |||| ||||||||||||||| || |||||||||||||||||| |||| ||||| |||||||||
Db 301 CATTAGTAGGCGGATCAGTATATATTGTGAGTGTGGCACATAAACGGCCATAAACGGATGTTG 360

Qy 629 ATTTTGGTGCTGAGGAGCAAAAATCCTGACAACATCGTTTACTTTACTTATAAAATTGTGAAAC 688
       |||||||||| ||||||||||||| ||||||||| || ||||||||||||||||||||||||||
Db 361 ATTTTGGTGCTGAGGAGCAAAAATCCTGACAACATCGTTTACTTTACTTATAAAATTGTGAAA 420

Qy 689 GGAATAATTATAA--------AAACGATCAAAACGCATCCTTATGAGAAAGACTACCCAACCCAC 745
       |||||||||| || ||| |||||||||| ||||||||||||||||| |||||||||||
Db 421 GGAATAATTATTCAGCCTTGGGAGGAGACAACATCCTTATGATGGAAGAATTACATATGGCCAC 480

Qy 746 GCTTACATAAAATTTGTTACGGAAGCCACCCCCAATCGATAATGACTTCTGATAATGAACGGCA 805
       |||||||| |||||||||||||| | | |||||||||||| |||||| || |||||| ||||||
Db 481 GTTTACATAAAATTTGTTACCGAAGCCGATAGGTATGA---CTATCCTGAGCGTGT 537

Qy 806 ACAAAATATACGATAGGACGAGAAAAATCCCGAACACGCGTGGGTATCGGTCCGGGTTGGCAGT 865
       |||||||||||||||||||||||||||| | ||||||||||||| ||||| |||||||||||||
Db 538 GCGTATAGGTTCAGGACACCAGTATTGTGCGGTACGGAGGTACGCAAGGATGA 582

Qy 866 TTTGGCGGAAACGGATCAAAAACAACGGCGACCAAAGTTGCCGCGCATAATCAAATTACTGACAG 925
       |||||| |||| ||||||||||||||||||||||||||||||| ||||||||||||| |||| |
Db 583 ----AGAAACAAAATGCCATATAGTTCATACGATAATTCAGAGCGCTATAACTACCTTTATG 638

Qy 926 CAGGCAAATCACACAACCAAAGGCGGCGGAGCAGGGGGCGGCGGCGGCTTGGTCGAAGTCGAGGCGATG 985
       |||| |||||||| |||| | ||||||||||| |||||||||||||||| |||||||||| ||||||
Db 639 CAGGAAATACACATCAAAGTAATGGCCATAAAGGTACAGTTCATTTTAGCCGGTAAGG 698

Qy 986 TGCGGCCAAGGGGGCAATTACGGCGCCCCCCATTCCTATTGCAGGCTCAGCGGCGGCGGCAGCGGTT 1045
       ||| ||||||||| |||||||||| |||||||||||||||||||| || ||||||||||||||||||
Db 756 CACCAATGTTTTATTTATTGATGCCGGAGAACAAAAATGGTTTATTAAAAACGGGGTAATTACAAA 815

Qy 1046 CGCCTATGTTTATTTATGATGCGGAAAAAACAAAAATGGTTGATTAAACGGCGGTTATTGAGGA 1105
        |||||| ||||||||||| |||||||||||||||||||||||||| |||||||| ||||||||
Db 699 TAATTAGACCTAAACCATTAAGGTCCATTAACCTTAT---AGGAGCTCAGGGTGATAGTGGCT 755

Qy 1106 CCGGCGAACCCTTTGGGCGGGGGGACAAGAGAATACATTCCAACTGGTACGGCAAGTCTTTTTTG 1165
        ||| |||| ||||||| |||||| ||| |||||| |||||||||||||||||| |||| |||||
Db 816 CTGGACATCCTTTGTGGGGAGAGGTAAATGGGTTTCAGCTAATACGTGAACGAATGGTTTT 875

Qy 1166 ATGAAATCCTTGAAAAAGATTGCCTACATCGTTTTATAGCCCATCGGGCAACTGGTGCAT 1225
        || |||| || ||||||| |||| ||||||| |||| ||||| ||||| |||||||||||
Db 876 ATACTGAAGTTTACCAG---------TTGAATACCCGTAGTGTTTTTCGACCGCT 920

Qy 1226 ACACCAATTACAGACAAAGGCGACGGCAGGGCCATTGTCAAACAACAAAACAGGAGGAAGACCAT 1285
        ||| |||| | |||||||| || |||||| |||||||||||||||||||||| |||||||||||
Db 921 ATATTCCCTCAAATAAATGGACATTATTTCCTTTGTATCAAATAATAATGATGGTTACAGGTAAAT 980

Qy 1286 CTGAAGTCCGGCCATCGGTTTAAAAGACGACAAAATTACCTGCCGAGGTAAAGACGAGGATGTTT 1345
        ||| ||||||| ||||||| |||| |||| ||| ||||||| ||||||||||| ||||| |
Db 981 TAACTTTAACTAGACCTAGTTAAAGAGATG---------------------GCTCTAAAGCTAAATC 1023

OM nucleic - nucleic search, using sw model

Copyright (c) 1993 - 2003 Compugen Ltd.

GenCore version 5.1.5

Run on: May 6, 2003, 00:47:41 ; Search time 930.919 Seconds
                                        (without alignments)
                                        1679.481 Million cell updates/sec

Title: PCT-US03-05226-16
Perfect score: 4828
Sequence: 1 tgaccgcaacttcgagaaa..........gactgcagtgtatacaattg 4828
Scoring table: IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0
Searched: 2185239 seqs, 112599159 residues
Total number of hits satisfying chosen parameters: 4370478
Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   N_Geneseq_101002:*
         1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
         2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
         3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
         4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
         5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
         6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
         7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
         8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
         9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
        10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
        11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
        12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
        13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
        14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
        15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
        16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
        17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
        18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
        19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
        20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
        21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
        22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
        23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
        24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	% Query Match	Length	DB ID	Description
c 1	3039	62.9	1830121	17	AAT42063	Haemophilus influe
c 2	2970	61.5	4319	17	AAT17215	Adhesion and penet
3	1177.6	24.4	4350	20	AAZ12252	Neisseria meningit
4	1108	22.9	4407	20	AAZ12253	Neisseria meningit
5	1083.4	22.4	4374	21	AAZ38918	Neisseria meningit
6	1081.8	22.4	4374	21	AAZ12251	Neisseria meningit
7	1081.8	22.4	4374	21	AAF21587	N. meningitidis pa
8	1081.8	22.4	4374	21	AAA81302	N. meningitidis pa
9	1081.8	22.4	26778	21	AAA81477	N. meningitidis pa
10	1081.8	22.4	349980	21	AAF21612	Neisseria meningit
11	1081.8	22.4	837096	21	AAA81499	N. meningitidis pa.
12	1041.2	21.6	4365	21	AAZ38917	Neisseria meningit
13	820.8	17.0	969	18	AAT88142	H. influenzae Hap
14	402.2	8.3	1312	21	AAA81400	N. meningitidis Me
15	401	8.3	2991	20	AAZ12250	Neisseria meningit
16	213.2	4.4	607	17	AAT26893	Haemophilus influe
17	193	4.0	5091	11	AAO06164	The Iga gene, enco
18	183.8	3.8	30078	21	AAA81520	N. meningitidis pa
19	183.8	3.8	349980	21	AAF21608	Neisseria meningit
20	183.8	3.8	1437668	21	AAA81690	N. meningitidis B
21	170	3.5	781	21	AAA81401	N. meningitidis Me
22	169.6	3.5	891	21	AAA81402	N. meningitidis Me
23	155.4	3.2	4296	21	AAZ54326	Neisseria meningit
24	155.4	3.2	16526	21	AAA81472	Neisseria meningit
25	155.4	3.2	172325	21	AAF21613	Neisseria meningit
26	155.4	3.2	349980	21	AAF21612	Neisseria meningit
27	155.4	3.2	837096	21	AAA81489	N. meningitidis pa
28	126.2	2.6	4113	22	AAB89172	Neisseria meningit
c 29	126.2	2.6	48254	22	ABA89141	Escherichia coli P
c 30	126.2	2.6	48345	22	ABA89142	Escherichia coli P
c 31	119.8	2.5	4128	22	ABA88994	Escherichia coli P
c 32	119.8	2.5	7654	22	ABA88991	Escherichia coli P
c 33	71.6	1.5	3399	17	AAT05868	Chicken leucocytoz
c 34	60.2	1.2	1686	16	AAQ87587	DNA encoding Leuco
c 35	56.6	1.2	642	21	AAA81729	N. meningitidis pa
c 36	56.4	1.2	7597	24	AAS46542	Tumour suppressor
37	56	1.2	10595	22	ABL33013	Escherichia coli P
38	55	1.1	580	21	AAA81832	N. meningitidis pa
c 39	54.8	1.1	12592	24	AAS61102	Human immune syste
c 40	54.6	1.1	5244	24	ABK40014	Human chemically p
41	54.2	1.1	3579	21	AAA70099	Plasmodium falcipa
c 42	53	1.1	10250	24	ABN80083	Human chemically m
43	53	1.1	11422	24	ABK39937	Human chemically p
44	53	1.1	11422	24	ABL32219	Human immune syste
45	52	1.1	7143	24	ABL32983	Human immune syste

ALIGNMENTS

RESULT 1
AAT42063/c
ID   AAT42063 standard; DNA; 1830121 BP.
XX
AC   AAT42063;
XX
DT   14-SEP-1999 (first entry)
XX
DE   Haemophilus influenzae complete genome sequence.
XX
KW   Genome; bacterium; Haemophilus influenzae; computer readable medium;
KW   expression modulating fragment; regulation; gene expression; vector;
KW   organism; open reading frame; ORF; ds.
XX
OS   Haemophilus influenzae.
XX
PN   WO9633276-A1.
XX
PD   24-OCT-1996.
XX
PF   22-APR-1996;   96WO-US05320.
XX
PR   07-JUN-1995;   95US-0487429.
PR   21-APR-1995;   95US-0426787.
PR   07-JUN-1995;   95US-0476102.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (UVJO) UNIV JOHNS HOPKINS.
XX
PI   Adams MD, Fleischmann RD, Smith HO, Venter JC, White O;
XX
DR   WPI; 1996-485782/48.

```
XX
PT  Haemophilus influenzae Rd genome recorded on computer readable
PT  medium - useful for identifying commercially important nucleic acid
PT  fragments by homology searching
XX
PN
PS
XX
CC  Claim 1: Page 77.2-77.1091; 1291pp; English.
CC
CC  This sequence represents the complete genome sequence of the bacterium
CC  Haemophilus influenzae strain Rd. The invention relates to a computer
CC  readable medium (CRM) having recorded upon it the complete H.influenzae
CC  nucleotide sequence (I), a representative fragment of (I) or a nucleotide
CC  sequence at least 99% identical to (I). By providing the full-length
CC  genomic sequence in a computer readable form, it is possible to identify
CC  commercially important nucleic acid fragments and expression modulating
CC  fragments (EMFs) of the Haemophilus genome. The EMFs can be used to
CC  regulate the expression of a nucleic acid molecule. Vectors and altered
CC  organisms comprising the predicted ORFS can be used to produce any of the
CC  polypeptide fragments of the H. influenzae Rd genome.
XX
SQ  Sequence 1830121 BP; 567399 A; 350615 C; 347389 G; 564036 T; 682 other;
```

```
Query Match     62.9%;  Score 3039;  DB 17;  Length 1830121;
Best Local Similarity   78.6%;  Pred. No. 0;
Matches 3891;  Conservative   2;  Mismatches 853;  Indels 202;  Gaps 14;
```

```
Qy      1  TGACCGCACTTTCAGAGAAAACTCACATAAAGTGCGGTTATTTATTTAGTGATATTGTTT   60
           ||||||| ||||||| ||||| ||||||||||||||||||||||||||||||||||||||
Db  278838  TGATCGCACTTTCAGAGAAAACTCACATAAAGTGCGGTTATTTATTTAGTGATATTGTTT  278779

Qy     61  TAATTTTAGTTATCTGTATAAATTACACATAATAATTCCATCGCCACAGTAAGATTACC  120
           |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  278778  TAATTTTAGTTATCTGTATAAATTACACATAATAATTCCATCGCCACAGTAAGATTACC  278719

Qy    121  CACTAAGTATTAAAGCAGAAAATTTGGCTTAATTATAATAATAAGTTTTTACTCAT     180
Db  278718  CACTAAGTATTAAGCCAAAACCTAGAAAATTTGGCTTAATTATAATAAGAATTACC    278659

Qy    181  TTATTTTTTTTTGTTGTCGCCTTTTAGTTCGTTTTTTTTAGCTGAAATCCCTTAGAAAATCACCG  240
Db  278658  TTATTTTCTTTTGTTCGGCCTTTAGTTCG-TTTTTTAGCTGAAATCCCTTAGAAAATCACCG     278600

Qy    241  CACTTTTATTGTTCAATAGTCGTTTAACCACGTAATTTTTTAATACGAAAAATTACTTAAT   300
Db  278599  CACTTTTATTGTTCAATAGTCGTTTAACCACGTAATTTTTTAATACGAAAAATTACTTAAT  278540

Qy    301  TAAAATAAACATTATGAAAAAAAACTGTAATTTCGTTTCGTTGAAATTTTTTAAACCGTT    360
Db  278539  TAAAATAAACATTATGAAAAAAAACTGTAATTTCGTTTCGTTGAAATTTTTTAAACCGTT   278480

Qy    361  TTAGGGATAGAGTAGTAGCTCAAGGTCACACTTCATACCTTGTGTTTCA          420
Db  278479  TTAGGGATAGAGTAGTAGCTCAAGGTCACACTTCATACCTTGTGTTTCA         278420

Qy    421  TATCGTGATTTTGCCGAGAATAAAGGGAGAAGTCACAGTTGGGGCTAAAATATATTGAGGTT  480
Db  278419  TATCGTGATTTTGCCAGAATAAAGGGAGAAGTCACAGTTGGGGCTAAAATATATTGAGGTT   278360

Qy    481  TACAATAAAAATTGAAATTTAGTTGGGCACATCAATGACAAAGCCCCAAGATTTGATTTT   540
Db  278359  TATAACAAAGAAGGGCCAATTAGTTGGGCACATCAATGACAAAGCCCCAAGATTTGATTTT  278300

Qy    541  TCCGTGGTTGTCGCGAAAATGGGTGGGCCATTGGTGGGCCATGGTGTTTTTTTTGAGTTGG   600
Db  278299  TCCGTGGTGTCGCGGAAAATGGGTGGGCCATTGGTGGGCCATGGTGTTTTTTTTGAGTTGG  278240

Qy    601  GCACATAATGTAGGCGTATACCAATCTGATTTTTGGTGGTCGGAAGCAAAAATCCTGATCAA  660
Db  278239  GCACATAACGGCGGATAATAACGAGTTGATTGGTGCAGAAGGACGAAACCTGATCAG     278180

Qy    661  CACGGCTTTACTTATCTAAAATTGTAAAAAGAAATTAATTAATATAATAA--AACGATCAAACGCAT  717
Db  278129  CACGGCTTTACTTATCAAATTGTAAAAAGAAATTAATTAATATAA--AACGATCAAACGCAT   278120
```

```
Qy    718  CCTTATGACAAAAGACTACCAACCCACGCTTACATAAAATTGTTACGGGAGCCACCCCA   777
Db  278119  CCTTATGATGGGGATTATCATAATGCCTCGTTAACATAAAATTGTAACTGAACATCCCGAA  278060

Qy    778  ATCGAATGACTTCTGATATGAAACGGCAACAAAATAACAGATAGGACGAAAAAATCCCGA  837
Db  278059  GTGGGTATGACACAGATATGGCAGTTTGGGGAAACGGATCAAAACAAG-----         278000

Qy    838  CGCGTGCGTGTATCGGCTCCGGGTGGCAGTTTGGGGGAAACGGATCAAAACAAG-----  889
Db  277999  CGTGTACGTGATAGGCTCAGAGCGTCAGTATTGGGGTACAGAAGAAAAGATGAGGAAT     277940

Qy    890  --------GCGACCAAGGTGCCGGGCGCATATACTTACCTGACAGCAGCAATACACAC  939
Db  277939  GTACATTAGTTCAATAAATTGTCTCAGTTGCATATGGCTTACTGACAGGAAAATACCAT    277880

Qy    940  AACCAAGGCGCGAGGGGGCGGCGACGCAGCCGTGGTTGCCAATAGTTGGCCCAAGCGGGC  999
Db  277879  ACTCAGATGGAAAATGGTTAAATGTACAGTGAATCTTAGTGGTAATGTAGTTAGCCCTAAT   277820

Qy   1000  AATTACGGCGCCCATTCCTATTGCAGCCTCAAGCGCGGCGACAGCCGGTTCGCCTATGTTT  1059
Db  277819  CATTAGGGTCCATTACCCAACCGGGTGGTTCTAAAGCGGATAAGCGGCCGGCAAAGTTTAT   277760

Qy   1060  TATGATGCGGCGAAAAACAAAATGGTTGAATAACGGGTATTTGAGGACGGCAACCCTTGG  1119
Db  277759  TATGATGGAGCAGAGAATACATTCCAACTGGTACG---CAAGTCTTTTTTTGATGAAATCCT   277700

Qy   1120  GCGGGACAGCAGAATACATTCCAAACTGGTACG---CAAGTCTTTTTTTGATGAAATCCTT  1176
Db  277699  TTCGGAAGAGGTAATGGGTTCAGTAATACGTGAAGAATGGCTTATATAAAGAAGTTCTT     277640

Qy   1177  GAAAAAGAGATTTGCGTGTACAATCGTTTT-----TATAGCCCATGGCGAATGGTCGATACACC  1230
Db  277639  GCGGTTGAATACCCCTAGTGTTTTCAACGGCTATATTCCCCAATAAATGACAACATTATTCC   277580

Qy   1231  ATTACAGACAAAGGCGGAGGCAGGGCAGCAGTTGTCA------------AACAA  1269
Db  277579  TTTGTATTCAAATAATGATGTTACAGGTAAATTAACTTTAACCTTAGTAAAGGGC   277520

Qy   1270  CAAACAGGAAGAGACCAATCTGAAGTCGCCATCGGGTTAAATTAATCTGCGGAACAAAAACC  1329
Db  277519  TCTAAAGCAAAATGAGGTAAGTACAGGTAAATTAACTTAACCATCGTTAAATCAAAAC     277460

Qy   1330  GGTAAAGACGGATGGTTTACCAAATACCAAGGTCCAAAAAATATACCGGCTTGGGAAAAAT  1389
Db  277459  GCTAAAGACAAACATGTTAGGCAGCGAACTGTGGAAGTCAAGAGAAAGGTATAGAAATAAC   277400

Qy   1390  GGTGGGAAACCTGTATTCGGATCAAAAAACGGCAGTCCAAAAATATATCCAACCAACATC  1449
Db  277399  GGAAAAATATTCCTTGGCGACCAAGAAAGGAACTTTAACAATCCGAAAATAATTATA      277340

Qy   1450  AACCAAGGTGCGGGGGTTTGTAATTGTAAGGTAACTTCACGGT---ACTTCAGAAAAT     1506
Db  277339  AATCAAGGTGCTGGGGATATTACTTCAGAGTTAATTTTGTGTAAAAGGCACAAAAAT    277280

Qy   1507  AATGCAACTTGGCAGGTGCTGAAGAACAACAGTACTTTACTTGTCAAGGGCAAGCAACATG  1566
Db  277279  GTTCACAAATCCTGAAAATGATCGTTTATCTAAAATAAGGCCACTTTATTAGTCAAT      277220

Qy   1567  GTAAATGGTGTTGAAAATGAATCGGCCTTCTTAAAATACGAAAGGCCACATTGCACGTTAA  1626
Db  277219  GTTCACAAATCCTGAAAATGATCGTTTATCTAAAATAAGGCCACTTTATTAGTCAAT      277160

Qy   1627  GCCAAAGGGGAAAATAAAGGTTGCATCAGCGTAGGGCCGGATAGGGGGATTAGTTTAAGTG  1686
Db  277159  GGTAAGGGGAAAAATTAGGAGTTTAAGTGCGGGCCGGATAGGGGGATTAGTTTAAGTG     277100

Qy   1687  CAGGCAGACGATCAAGCAACAAACAACAAGCCTTAGTGCGGGCTTGGTTAGTTGGGACCAG  1746
Db  277099  CAAGCAGATGAAGCGGGTCAAAAAACAACAGTTTCAAAGAAGTTTCGAAGCGGTCGAAGA   277040

Qy   1747  GGTACGGTTCAGTTAAACGATGACAGAGCAAATTTAATACTGATAAAATTTTATTCGGCTTC  1806
Db  278129  CACCGCTTTACTTATCAAATTGTAAAAAGAAATTAATTAATATAATAATAA      277040
```

```
          |  || ||||| |||||   |    | ||| || || || | ||  | |||||||| ||
Db 277039 GCAACAGTTCAATTAAATAGTACAGATCAAGTTGATCCTAACAATATCTATTTCGGATTT 276980

Qy   1807 CGTGGTGGTCGCTTAGATCTTAATGGGCATTCATTAACCTTTAAACGTATCCAAAATACG 1866
          ||||||||||||||||||||||| |||||||||||||||||||| |||||||||||||||
Db 276979 CGTGGTGGTCGCTTAGATCTTAACGGGCATTCATTAACCTTTAAGCGTATCCAAAATACG 276920

Qy   1867 GATGAGGGAGCAACGATTGTTAATCACAATGCCACAACAGAATCTACAGTGACCATTACT 1926
          || ||||| || | |||||| || || ||| | ||   ||   |||   | || ||||||
Db 276919 GACGAGGGCGCGATGATTGTGAACCATAATACAACTCAAGTCGCTAATATTACTATTACT 276860

Qy   1927 GGCAGCGATACCATTAATG---------------ACAACACTGGCGATTTAACCAATAAA 1971
          || | ||| |  |||| ||               | || | |    |  |     | |
Db 276859 GGGAACGAAAGTATTACTGCTCCATCTAATAAAAAGAATATTAATAAACTTGATTACAGC 276800

Qy   1972 CGTGATATTGCTTTTAATGGTTGGTTTGGTGATAAAGATGATACTAAAAATACTGGACGT 2031
             || ||||| |  || ||||||||||| || | |||| | | |||| | | ||||||
Db 276799 AAAGAAATTGCCTACAACGGTTGGTTTGGCGAAACAGATAAAAATAAACACAATGGACGA 276740

Qy   2032 TTGAATGTTACTTACAATCCGCTTAACAAAGATAATCACTTCCTTCTATCAGGTGGAACA 2091
          || ||  ||| ||| || ||    |   |||||  |   || || || |||||||| |||
Db 276739 TTAAACCTTATTTATAAACCAACCACAGAAGATCGTACTTTGCTACTTTCAGGTGGCACA 276680

Qy   2092 AATTTAAAAGGCAATATTACTCAAGACGGTGGCACTTTAGTGTTTAGTGGTCGCCCAACA 2151
          || ||||||||| ||||||| |||      || |   || | || || ||| | || |||
Db 276679 AACTTAAAAGGCGATATTACCCAAACAAAAGGTAAACTATTTTTCAGCGGTAGACCGACA 276620

Qy   2152 CCACACGCATACAATCATTTAAA------TCGCCTAAACGAGCTTGGGCGACCTAAGGGC 2205
          || ||||| |||||||||||| |      | |   | |   |   ||    ||  | |||
Db 276619 CCCCACGCCTACAATCATTTAGACAAACGTTGGTCAGAAATGGAAGGTATCCCACAAGGC 276560

Qy   2206 GAAGTGGTTATTGATGACGATTGGATCAACCGTACATTTAAAGCTGAAAACTTCCAAATT 2265
          ||| | ||    ||| ||||||||:|||||||||||||||||||||||||||||||||||
Db 276559 GAAATTGTGTGGGATTACGATTGGWTCAACCGTACATTTAAAGCTGAAAACTTCCAAATT 276500

Qy   2266 AAAGGCGGAAGTACGGTGGTTTCTCGCAATGTTTCTTCAATTGAAGGAAATTGGACAATC 2325
          |||||||||||| ||||||||||||||||||||||||||||||| |||||||||||| ||
Db 276499 AAAGGCGGAAGTGCGGTGGTTTCTCGCAATGTTTCTTCAATTGAGGGAAATTGGACAGTC 276440

Qy   2326 AGCAATAACGCCAACGCGACATTTGGTGTTGTGCCAAATCAACAAAATACCATTTGCACG 2385
          |||||||| || || || ||||||||||||||||||||| ||||||||||||||||||||
Db 276439 AGCAATAATGCAAATGCCACATTTGGTGTTGTGCCAAATTAACAAAATACCATTTGCACG 276380

Qy   2386 CGTTCAGATTGGACAGGATTAACGACTTGTAAAACAGTTAATTTAACCGATAAAAAAGTT 2445
          |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db 276379 CGTTCAGATTGGACAGGATTAACGACTTGTAAAACAGTTAATTTAACCGATACAAAAGTT 276320

Qy   2446 ATTGATTCCATACCGACAACACAAATTAATGGCTCTATTAATTTAACTAATAATGCAACA 2505
          ||| |||||||||| | |||||||||||||||||||||||||||||||||||||||||||
Db 276319 ATTAATTCCATACCCATAACACAAATTAATGGCTCTATTAATTTAACTAATAATGCAACA 276260

Qy   2506 GTGAATATTCATGGTTTAGCAAAACTTAATGGTAATGTCACTTTAATAAATCATAGCCAA 2565
          |||||||||||||||||||||||||||||||||||||||||||||||| |||| ||||||
Db 276259 GTGAATATTCATGGTTTAGCAAAACTTAATGGTAATGTCACTTTAATAGATCACAGCCAA 276200

Qy   2566 TTTACATTGAGCAACAATGCCACCCAAACAGGCAATATCCAACTTTCAAATCACGCAAAT 2625
          ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db 276199 TTTACATTGAGCAACAATGCCACCCAAACAGGCAATATCAAACTTTCAAATCACGCAAAT 276140

Qy   2626 GCAACGGTGGATAATGCAAACTTGAACGGTAATGTGCATTTAACGGATTCTGCTCAATTT 2685
          ||||||||  ||||||| |  || ||||| ||||||||||||||||||||||||||||||
Db 276139 GCAACGGTAAATAATGCCACGTTAAACGGCAATGTGCATTTAACGGATTCTGCTCAATTT 276080

Qy   2686 TCTTTAAAAAACAGCCATTTTTCGCACCAAATTCAGGGCGACAAAGACACAACAGTGACG 2745
          |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db 276079 TCTTTAAAAAACAGCCATTTTTGGCACCAAATTCAGGGCGACAAAGACACAACAGTGACG 276020

Qy   2746 TTGGAAAATGCGACTTGGACAATGCCTAGCGATACTACATTGCAGAATTTAACGCTAAAT 2805
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 276019 TTGGAAAATGCGACTTGGACAATGCCTAGCGATACTACATTGCAGAATTTAACGCTAAAT 275960

Qy   2806 AATAGTACTGTTACGTTAAATTCAGCTTATTCAGCTAGCTCAAATAATGCGCCACGTCAC 2865
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 275959 AATAGTACTGTTACGTTAAATTCAGCTTATTCAGCTAGCTCAAATAATGCGCCACGTCAC 275900

Qy   2866 CGCCGTTCATTAGAGACGGAAACAACGCCAACATCGGAAGAACATCGTTTCAACACATTG 2925
          ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db 275899 CGCCGTTCATTAGAGACGGAAACAACGCCAACATCGGCAGAACATCGTTTCAACACATTG 275840

Qy   2926 ACAGTAAATGGTAAATTGAGCGGGCAAGGCACATTCCAATTTACTTCATCTTTATTTGGC 2985
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 275839 ACAGTAAATGGTAAATTGAGCGGGCAAGGCACATTCCAATTTACTTCATCTTTATTTGGC 275780

Qy   2986 TATAAAAGCGATAAAATAAAATTATCTAATGACGCTGAAGGCGATTACACATTAGCTGTT 3045
          ||||||||||||||| |||||||||| ||||||||||| ||||||||||||||| |||||
Db 275779 TATAAAAGCGATAAATTAAAATTATCCAATGACGCTGAGGGCGATTACACATTATCTGTT 275720

Qy   3046 CGCGACACAGGCAAAGAACCTGTGACCCTTGAGCAATTAACTTTAATTGAAGGCTTGGAT 3105
          ||| |||||||||||||||||||||||||||||||||||||||| |||||| |||| |||
Db 275719 CGCAACACAGGCAAAGAACCTGTGACCCTTGAGCAATTAACTTTGATTGAAAGCTTAGAT 275660

Qy   3106 AATCAACCCTTGCCAGATAAGCTAAAAATTACTTTAAAAAATAAACACGTTGATGCGGGT 3165
          ||| |||| ||  |||||||||| ||| |||||||| ||||| | ||||||||||| |||
Db 275659 AATAAACCGTTATCAGATAAGCTCAAATTTACTTTAGAAAATGACCACGTTGATGCAGGT 275600

Qy   3166 GCATGGCGTTATGAATTAGTGAAGAAAAACGGCGAATTCCGCTTGCATAATCCAATAAAA 3225
          ||||  |||||| ||||||||||||| || |||||||||||||||||||| |||||||||
Db 275599 GCATTACGTTATAAATTAGTGAAGAATAAGGGCGAATTCCGCTTGCATAACCCAATAAAA 275540

Qy   3226 GAGCAGGAATTGCGCAATGATTTAGTAAAAGCAGAGCAAGTAGAACGAGCATTAGAAGCA 3285
          ||||||||||||| |||||||||||||| ||||||||||| ||||| | ||||||||||
Db 275539 GAGCAGGAATTGCTCAATGATTTAGTAAGAGCAGAGCAAGCAGAACAAACATTAGAAGCC 275480

Qy   3286 AAACAAGCTGAACTGACTACTAAAAAACAAAAAACTGAGGCTAAAGTGCGGTCAAAAAGA 3345
          ||||||| ||||| |||| || |||||||||||| | |||| |||| |||||||| ||||
Db 275479 AAACAAGTTGAACAGACTGCTGAAAAACAAAAAAGTAAGGCAAAAGCGCGGTCAAGAAGA 275420

Qy   3346 GCGGCGTTTTCTGATACCCCGCCTGATCAAAGCCAGTTAAACGCATTACAAGCCGAACTC 3405
          |||| ||| |||||||||||| ||| |||||||| ||||||||||||| ||||| |||
Db 275419 GCGGTGTTGTCTGATACCCCGTCTGCTCAAAGCCTGTTAAACGCATTAGAAGCCAAAC-- 275362

Qy   3406 GAGACGATTAATGCCCAACAGCAAGTGGCACAAGCGGTGCAAAATCAGAAAGTAACTGCA 3465
             | | | ||    | || | ||     |||||     ||||| || |||      |
Db 275361 ---AAGTTGAACAGACTACTGAAACACAAACAAGTAAGCCAAAAACAAAAAAAGGGCGGT 275305

Qy   3466 CTTAACCAAAAGAACGAGCAAGTTAAAACCACTCAAGATAAAGCAAATTTAGTCTTGGCA 3525
          |  ||  |  |    | |||
Db 275304 CAAAAAGAGCATTGAGTGCA---------------------------------------- 275285

Qy   3526 ACTGCATTGGTGGAAAAAGAAACCGCTCAGATTGATTTTGCTAATGCAAAATTAGCTCAG 3585
                                           |  ||| || || |    || | ||
Db 275284 ---------------------------------GCGTTTTCTGATACCCCGTTTGATCTA 275258

Qy   3586 TTGAATTTAACACAACAACTAGAAAAAGCCTTAGCAGTGGCTGAGCAAGCAGAAAAAGAG 3645
              | ||||    |     ||  |||      |   |   ||  ||| |  ||    |
Db 275257 AGCCAGTTAAAGGTATTCGAAGTCAAACTTGAGGTTATTAATGCCCAACCGCAAGTGAAA 275198

Qy   3646 CGTAAAGCTCAAGAGCAAGCGAAAAGACAACGCAAACAAAAAGACTTGATCAGCCGTTAT 3705
              || ||||||| |||| | |   |||| ||||||||||||| ||||||||||||||
Db 275197 AAAGAACCTCAAGATCAAGAGGA---ACAAGGCAAACAAAAAGAATTGATCAGCCGTTAC 275141

Qy   3706 TCAAATAGTGCGTTATCAGAATTATCTGCAACAGTAAATAGTATGCTTTCCGTTCAAGAT 3765
          ||||||||||||||||| || || ||||||||||||||||||||| ||||||||||||||
Db 275140 TCAAATAGTGCGTTATCGGAGTTGTCTGCAACAGTAAATAGTATGTTTTCCGTTCAAGAT 275081

Qy   3766 GAATTAGATCGTCTTTTTGTAGATCAAGCTCAATCTGCGGTGTGGACAAATATCTCACAG 3825
          ||||| ||||||||||||||||||||||| ||||||||  |||||||||||||| |||||
Db 275080 GAATTGGATCGTCTTTTTGTAGATCAAGCACAATCTGCCCTGTGGACAAATATCGCACAG 275021

Qy   3826 GATAAAAGACGTTATGATTCTGATGCGTTCCGTGCTTATCAGCAGAAAACGAACTTGCGT 3885
          ||||||||||| ||||||||||||||||||||||||||||||||||||||| ||||||||
Db 275020 GATAAAAGACGCTATGATTCTGATGCGTTCCGTGCTTATCAGCAGAAAACGNACTTGCGT 274961

Qy   3886 CAAATTGGGGTGCAAAAAGCCTTAGCTAACGGACGAATTGGGGCAGTTTTCTCGCATAGC 3945
          ||||||||||||||||||||||||| ||| |||||||||||||| |||||||||||||||
Db 274960 CAAATTGGGGTGCAAAAAGCCTTAGATAATGGACGAATTGGGGCGGTTTTCTCGCATAGC 274901
```

```
Qy    3946 CGTTCAGATAATACTTTTGATGAATGAACAGGTTGTAAAAATCACGCAACATTAACGATGATGTCG
           ||||||||||||||||||||| ||||||||||||||  |||||||||||| ||||||||||||||||
Db  274900 CGTTCAGATAATACTTTTGACGAATGAACAGGTTAAAAATCACGCACATTAACGATGATGTCG  274841

Qy    4006 GGTTTTGCCCAATATCAAATGGGGTGATTTACAATTTGGTGTAAAAATCACGGCGAACGGGAATT
           ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db  274840 GGTTTGCCCAATATCAAATGGGGCGATTTACAATTTGGTGTAAAAATCACGGCGAACGGGAATT  274781

Qy    4066 AGTGCGAGTAAAAATGGCTGAGAGAACAAAGCCGAAAAAATTCATCGAAAAGGCGATAAATTAT
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  274780 AGTGCGAGTAAAAATGGCTGAGAGAACAAAGCCGAAAAAATTCATCGAAAAGGCGATAAATTAT  274721

Qy    4126 GGCGTGAAATGCAAAGTTATTCGTTCCATTTAGGGCAAATGGGTATTCAGCCTTATTTGGA
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  274720 GGAGTGAATGCAAGTTATCAGTTCCGTTTAGGGCAAATGGGTATTCAGCCTTATTTGGGT  274661

Qy    4186 GTTTAATCGCTATTTTATTGAAGGTAAAAAATTATCAATCTGAGGAAGTGAAAGTGCAAACA
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  274660 GTTTAATCGATATTTTATTGAACGTGAAAAATTATCAATCTGAAGAAGTGAAAGTGCAAACA  274601

Qy    4246 CCGAGCCCTTGCAATTTAATCGCTATAAATGCTGAGTACGGGTCGATTATTACGTTTACCCCG
           ||||||||||||||||||||| || || ||||||||||||||||||| ||||||||||||||
Db  274600 CCGAGCCTTGCATTTAATCGCTATAAATGCTGAGTACGGGTCGATTATTACATTTACCCCG  274541

Qy    4306 ACAGAGAAATATCAGCGTTAAGCGTTAAGCCTTATTTCTTCGTCAATTATAGTTGATGTTTCAAACGCT
           |||||||||||||||||||||||||||||||| || ||||||||||||||||||||||||||
Db  274540 ACAAATAATATCAGCGTTAAGCCTTATTTCTTCGTCAATTATAGTTGATGTTTCAAACGCT  274481

Qy    4366 AACGTACAAACCACTGTAAATCGGCGCGGGTGTTGCACAACACCATTGGACGGTTATTGGCAA
           ||||||||||||||||||||||||||||||||||||| || ||||||||||||||||||||
Db  274480 AACGTACAAACCACTGTAAATAGCACGCACGATGTTGCACAACACCATTGTTGATGTTTCAAACGCT  274421

Qy    4426 AAAGAAGTGGGATTAAAAGCGGGAAAATTTTACATTCCAACTTTCTGCTTTATTTCTAAA
           |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db  274420 AAAGAAGTGGGATTAAAAGCAGCAGAAAATTTACATTCCAACTTTCTGCTTTATTTCTAAA  274361

Qy    4486 TCTCAAGGTTCGCCAACTCGGTAAACACGCGGTGGAAATTAGGATATCGTTGG
           |||||||||||||||||||||||| || ||||||||||||||||||||| ||
Db  274360 TCTCAAGGTTCGGCAACTCGCAAAACAGCAAAATGTGGGCGGTGAAAATTGGGGTATCGTTGG  274301

Qy    4546 TAAAAATCAACATAA
           ||||||||||||||    TT
Db  274300 TAAAAATCAACATAA---------------TT  274241

Qy    4563 TTATTCTAATAATGGAACTTTATTTATTTAATTAAAAGTATCTAAGTAGTAGCACCC--TATAGGGG
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  274240 TTATTTCAATAATGGAACTTTATTTAATTAAGAGCACTCAAGTAGCACCCATAATAGGGG  274181

Qy    4621 ATTAATTAAGAGAGGATTTAATAATGAATTAACTAAAATTTACCGGATTGCTGCTGCA
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  274180 ATTAATTAAGAGGATTTAATAATGAATTAACTTTACCGGATTGCTGCTGCA  274121

Qy    4681 GTCGTATTATCTGCTTGTGCAAAGGATGCACCTGAAAGTGACAAAATCATCTCGGCGAAATA
           |||| |||||||||||||||||||||||||||||||||||||||||| |||||||||| |||
Db  274120 GTCGTATTATCTGCTTGTGCAAAGGATGCACCTGAAAGTGACAAAATCATCTCGGCAAATA  274061

Qy    4741 GCTGAAAGCAAACACTTCCAACATACTGATAAAACAAGTTGTATATTCTTGCAATTAG
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  274060 GCTGAAATGCAAACACTTCCAACACTGATAAAACACTGATAAAATATCCTGCAATAA  274001

Qy    4801 CAAACTGTGACTCCAGTGTATCAATTTG
           |||||||||||||||||||||||||||
Db  274000 CAAACGGTGACTGTGTGTATCAATTTG  273973
```

XX
DE   Adhesion and penetration protein gene.
XX
KW   Adhesion and penetration protein; hap gene; promoter; terminator;
KW   inverted repeat; stem-loop; protease; outer membrane protein;
KW   beta-domain; secretion; recombinant vaccine; monoclonal antibody;
KW   diagnostic; immunoassay; ds.
XX
OS   Haemophilus influenzae.
XX

FH Key	Location/Qualifiers
FT -35_signal	5..10
FT	/*tag= a
FT -10_signal	26..31
FT	/*tag= b
FT sig_peptide	60..134
FT	/*tag= c
FT CDS	60..4244
FT	/*tag= d
FT	/product= Adhesion and penetration preprotein
FT mat_peptide	135..4241
FT	/*tag= e
FT	/product= Adhesion and penetration protein
FT mat_peptide	135..2936
FT	/*tag= f
FT	/product= Secreted 110-kDa protease fragment
FT mat_peptide	2937..4241
FT	/*tag= g
FT	/product= 45-kDa outer membrane protein fragment
FT terminator	4284..4302
FT	/*tag= h
FT repeat_region	4284..4302
FT	/*tag= i
FT	/rpt_type= INVERTED
FT stem_loop	4284..4302
FT	/*tag= j

XX
PN   WO9605858-A1.
XX
PD   29-FEB-1996.
XX
PF   16-AUG-1995; 95WO-US010661.
XX
PR   25-AUG-1994; 94US-0296791.
XX
PA   (STRD) UNIV LELAND STANFORD JUNIOR.
PA   (UNIW) UNIV WASHINGTON.
XX
PI   Falkow S, St Geme JW;
XX
DR   WPI; 1996-151147/15.
DR   P-PSDB; AAR92768.
XX
PS   Claim 5; Fig 6; 105pp; English.
XX
PT   Haemophilus adhesion and penetration protein and corresponding DNA
PT   - used to produce vaccines against H. influenzae infection
XX
CC   The sequence encodes a Haemophilus influenzae adhesion and
CC   penetration protein. The sequence (hap gene) includes putative
CC   -10 and -30 sequences and a putative rho-independent terminator 3'
CC   to the hap stop codon. The terminator contains interrupted
CC   inverted repeats, with the potential for forming a hairpin
CC   structure containing a loop of 3 bases and a stem of 8 bases,
CC   followed by a stretch rich in T residues. The gene product is
CC   first synthesised as a preprotein, which is transported to the
CC   periplasm, followed by insertion of the C-terminal beta-domain into
CC   the outer membrane, possibly forming a pore, and export of the
CC   N-terminal fragment through the outer membrane, followed by
CC   autoproteolytic cleavage and secretion of the mature protease, leaving
CC   an outer membrane protein fragment. The gene may be inserted in a
CC   vector and expressed in recombinant host cells, for use as a
CC   recombinant vaccine. The gene product may also be used in

```
CC  diagnostic monoclonal antibody production.
XX
SQ  Sequence 4319 BP; 1497 A; 776 C; 891 G; 1155 T; 0 other;

Query Match         61.5%;  Score 2970;  DB 17;  Length 4319;
Best Local Similarity   82.3%;  Pred. No. 0;
Matches 3581;  Conservative   0;  Mismatches 645;  Indels 124;  Gaps 10;
```

```
Qy    253 TCAATAGTCGTTAACCACGTATATTTATTTTTTAATACGAAAAATTACTTAATTAAATAAACATT  312
          ||||||||||||||||||||||||||||| | |||||||||||||||||||||||||||||||||
Db      1 TCAATAGTCGTTAACTA-GTATTTTTTATATATTTTTAATACGAAAAATTACTTAATTAAAACATT   59

Qy    313 ATGAAAAAAACTGTATTTCGTCTGAATTTTTATTTTTTTTATTAGGGATAGTA  372
Db     60 ATGAAAAAAACTGTATTTCGTCTTAATTTTTAACCGCTTGCATTTCATTAGGGATAGTA  119

Qy    373 TCGCAAGCCGTGGGCAGGTCATACATTATTTTGGGATTGACTACCAATATATATCGTGTATTTT  432
Db    120 TCGCAAGCCGTGGGCTGGTCACACTTATTTTGGGATTGATTATTATTACCAATATTATCGTGATTTT  179

Qy    433 GCCGAGAAATAAAAGGGAAGTTCACAGTTGGGGTTTTACAATAAAAT  492
Db    180 GCCGAGAATAAAAGGGAAGTTCACAGTTGGGGCTCAAAATAATTATAGTTTATTAACAAACAA  239

Qy    493 GGAAAATTACTTGGCACATCAAATGACAAAAGCCCCCAAATGATTGATTTTTTCCGTGGTTGTCG  552
Db    240 GGGCAATTAGTTGGCACAATCAAATGACAAAAGCCCCGATAGTTGATTTTTTCTGTAGTGTCA  299

Qy    553 CGAAAATGGGGTGGGCATTGGGTGGGCGGATCAGTGTATATATIGTTGGGTGGGCACATAATGTA  612
Db    300 CGTAACGGCGTGGGCAGCCTTGGTTGGAAATCAATATTATTGTGAGGGTGGGCACATAACGTA  359

Qy    613 GGCTATACCAAATGTGGGATTTTGGTGTGCTGAAGCAACAAAATCCTGATCAACAATCGTTTTACT  672
Db    360 GGAATAACAGATGTTGATTGGTTGCAGAGGGAAAACAACCCCGATCACACAATCGTTTTACT  419

Qy    673 TATAAAAATTGTGAAACGGAATAAATTATTAAAAGCATCCTTTATGAGAAAGAC  732
Db    420 TATAAGATTGTAAAACGATAAAAACTACAAAAAAAAGATTAATTACATCGTTATAGAGGACGAT  479

Qy    733 TACCACAAACCCACGCTTACATCAAATTGTTACGGGAAGCACACCCCAAATCGATATTGACTTCT  792
Db    480 TACCATAATCCACGATTACATAAATTCGTTATGACGAAGCGCTCCAATTGATATATGACTTCG  539

Qy    793 GATATGAAACGGCAACAAAATAACAGATAGGACGGAAAATAATCCCGGACACGCGTCCGTATCGGC  852
Db    540 AATAATGAAATGGCCAGTACTTATTCAGATAGAACACAAAATAATCCAGAACGTGTTCGTATCGGC  599

Qy    853 TCCGGGTGTGGCAGTTTTGGGCGAAAACGAATCAAAAACAACACGGGGACCCAAGTTGCCCGGCCGGCCATAT  912
Db    600 TCTGGACGGCAGTTTTGGCGAAATGATCACACACAAGGGCATATGAAATTGGGGATTATICGTAT  659

Qy    913 CATTACCTGACAGCAGGCAATATCAACACACAACCAAGGGGAGAGCAGGGGGCGGCGGGTTTATT  972
Db    660 CATTAATCTGACAGCTGGCAATAAACAATCACAAAGCGACAAGTTGCCGGTGCAAGT  719

Qy    973 CTGAGCGGCGGATIGTGCCCAAGCGGCGGCAATTACGGCGCCCATTCCTATTGCAGGCTCAAGC  1032
Db    720 TTGGGAGGCGGATIGTTCGTAAAACGGGAGAAATATIGGTCCATTACCGATTGCAGGCTCAAAG  779

Qy   1033 GGCGCAGCGCGGTTCGCCCTATATTTTTATTATTATGATGCCGGGAAAAATTGGTGTTGATTAAC  1092
Db    780 GGGACAGTGGTTCTCCGGATGTTTATTTTATGTAGTGCTGAAAAACAAAAAAATGGTTAATTAAT  839

Qy   1093 GGCGTATTGAGGACCGGCCAACCCCTTGGGGCGGGACAGAGAATACAATTCCAACTGGTACGC  1152
Db    840 GGGAATTACGGGAAGCCAACCCTTTTGAAGGCAAAGAAAAAATGGGGTTTCAATTGGTTCGC  899

Qy   1153 AAGTCTTTTTTTTGAATGAAAATCCTGAAAAAGATTGGGCTACTACGTTTTATATAGCCCATCG  1212
Db    900 AAATCTTTATTTGATTGAAATTTTCGAAAAGAGAGAATTTACATACAATCCACTTACACCCGAGCT  959

Qy   1213 GCCAATGGTGCATACACCATTACGACGACAAAAGCGCGACCGGCGCAGCGCCCATTGTCAAACAACA  1272
Db    960 GGTAATGGAGTGTACACAATTAGTGGACAATTAGGGAAATGATAAATGGTCAGGGGGTCGTATAACTCAGAA  1019

Qy   1273 ACAGGAAGAGACCAGCTGGAGTGCACAATTAGTGGGAAATGATGAAATGGTCAGGGGGTCGTATAACTCAGAA  1332
Db   1020 TCAGGAATACCATCAGAAATTAAAATTACGTTAGCGAAAATATGAGTTTACCTTTGAAAGAG  1079

Qy   1333 AAAGACGATGTTTACCA-------ATACCAAGGTCCAAAATATATACCTCGCTTTTGAAT  1386
Db   1080 AAGGATAAAGTTCATAAATCCTAGATAATGACGGACCTAAATICATTTCGCCATCTGAC  1139

Qy   1387 AACGGTGGAAACCTTGTATTTCGGGAGATCAAAAAAACGGCACCTGTTTAAAGGACGAAGGT  1446
Db   1140 AATGGAGAAACGCTATAATTTTATGGATCAAAAAACAAGGCACCTGTTTAAAGGACGAAGGT  1199

Qy   1447 ATCAACCAAGGTGCCGGCGGCGTTTTGTATTTGAGGGTAACTTTACGGTATCTCAGAAAAT  1506
Db   1200 ATTAACCAAGGGCGGCGGTGGTCTTTATTTGTGAGGGTAATTTACAGTATGCTCCAAAATTCT  1259

Qy   1507 AATGCAACTTGGCAAGGTGCTGGAGTGCATGTAGGTGAAGACAGTACTGGTTACTTGGGAAA  1566
Db   1260 AACCAAACTTGGCAAGGAGCTGGCATACATGTAAGTGAAAATAGCACCGTTACTTGGGGAAA  1319

Qy   1567 GTAAATGGTGTTGAAAATGATCGGCCTTTTCTAAAAATCGGGTAGCGGAAAGGCCACGTTAAA  1626
Db   1320 GTAAATGGCGTTGGAACATGAATCGACTTTCTAAAAATTGGTAAAGGGAACCATTGCACGTTCAA  1379

Qy   1627 GCCAAAGGGGAAAAATAAAAGTTCAGCGGGTAAGTGGTAAAGTCATTTTGGAGCAG  1686
Db   1380 GCCAAAAGGGGAAAAATAAAAGTTCGATCAACGGTAGGCGGATGGGTAAAGTCATTTTGGAGCAG  1439

Qy   1687 CAGGCGAGACGATCAAAGGCAACAAACAAAGCCTTTAGTGAAATTGGCTTGGTTAGTGGGCAGA  1746
Db   1440 CAGGCGACGATCAAAGGCAACAAACAAAGCCTTTACTGAAATTGGCTTGGTTAGTGGGCAGA  1499

Qy   1747 GGTACGGTTCAGTTAAAACGATGACAAGCAATTTAAATACTGAATAAAATTTTATTTCGGGCTTC  1806
Db   1500 GGGACTGTTCAATTAAACGATAAAACAATTTGATAAACCGGATAAAATTTTATTTCGGGTTTT  1559

Qy   1807 CGTGGGTGGTCGCCTTAGATCTTAAATGGGCATTCATTCAATTTTAAACGTATCGGCAAAAATACG  1866
Db   1560 CGTGGGTGGTCGCCTTAGATCTTAAACGGGCATTCATTAAACGTATCCAAAATACG  1619

Qy   1867 GATGAGGGAGCAACGACGCATTTGTTTAATCACAAATGGGCCACAACAGAATTTACGGACCATTACT  1926
Db   1620 GACGAGGGGGCGGCAATGATTGGAACCATAATAACACTCAAGCGCGTAAATGTCACTATTACT  1679

Qy   1927 GGCACGGATACCATTAATGACAACACTGGGCGATTTAACCAATAAACGTGAT---------  1977
Db   1680 GGGAACGAAAAACATTGTTCTACCTAATGGGAAATAAATATTAATTACAGAAAA  1739

Qy   1978 ---ATTGCTTTAAIGGTTGGTTGGTTGATAAAGGATGAAAATACTAAAAAATACTTGGACGTTTG  2034
Db   1740 GAAATTTGCCTACACCGGTTTGGTTGCGTTTGGCGAAAATAAAACAAATGGGGGATTA  1799

Qy   2035 AAAIGTTACTTACAATCCGCTTAACAAAGATAAATCAACTTCCTTCTATCGGTACCGAACAAAT  2094
Db   1800 AACCTTATTTATAAAACCAACACAGAGATCGTTACTTGCTAACTTCAGGTTGGTAACAAAT  1859

Qy   2095 TTAAAAAGGCAATAATTACTCAGACACGGTGGCACTTTAGTGTTAGTGGTTCGCGGCAAACACCA  2154
Db   1860 TTAAAAGGCGGATAATTACCCAAACAAAAGGTAAACTATTTTTCAGCGGGTTAGTGACCCACG  1919

Qy   2155 CACGCATACACTAICAAICATTTTAAA------TCGCCTAAACGAGCTTTGGGCGACCTTA  2208
Db   1920 CACGCCTTACAAICAICATTTAAATAAAACGTTGGTCAGAAATGGAAGGTTATACCACAGGGCGAA  1979

Qy   2209 GTGGTTATTTGATTGAACGACATTTAAAGCTGAAAACTTCCAAATTAAA  2268
Db   1980 ATTGTGTGGGGATTCACCGATTTGATCAACCGTTACATTTTAAAGCTGAAAAACTTCCAAATTAAA  2039

Qy   2269 GGCGGAAGTACGGTGGTTCTCGCAATGTTCTTCGGAAGGAAATTGGACAAGAATCAGC  2328
Db   1980 (cont'd)
```

This page is a DNA sequence alignment printout (Qy / Db query–database rows with flanking position numbers).

Left column:

Row	Start	End
Db	2040	2099
Qy	2329	2388
Db	2100	2159
Qy	2389	2448
Db	2160	2219
Qy	2449	2508
Db	2220	2279
Qy	2509	2568
Db	2280	2339
Qy	2569	2628
Db	2340	2399
Qy	2629	2688
Db	2400	2459
Qy	2689	2748
Db	2460	2519
Qy	2749	2808
Db	2520	2579
Qy	2809	2868
Db	2580	2636
Qy	2869	2928
Db	2637	2696
Qy	2929	2988
Db	2697	2756
Qy	2989	3048
Db	2757	2816
Qy	3049	3108
Db	2817	2876
Qy	3109	3168
Db	2877	2936
Qy	3169	3228
Db	2937	2996
Qy	3229	3288
Db	2997	3056
Qy	3289	3342
Db	3057	3116
Qy	3343	3399
Db	3343	3176

Right column:

Row	Start	End
Qy	3400	3459
Db	3177	3231
Qy	3460	3519
Db	3232	3287
Qy	3520	3579
Db	3288	3321
Qy	3580	3639
Db	3322	3332
Qy	3640	3699
Db	3333	3392
Qy	3700	3759
Db	3393	3452
Qy	3760	3819
Db	3453	3512
Qy	3820	3876
Db	3513	3572
Qy	3877	3936
Db	3573	3632
Qy	3937	3996
Db	3633	3692
Qy	3997	4056
Db	3693	3752
Qy	4057	4116
Db	3753	3812
Qy	4117	4176
Db	3813	3872
Qy	4177	4236
Db	3873	3932
Qy	4237	4296
Db	3933	3992
Qy	4297	4356
Db	3993	4052
Qy	4357	4416
Db	4053	4112
Qy	4417	4476
Db	4113	4172

```
Qy   4477 ATTTCTAAATTCAAGGTTGGCACTCGGTAAAAGAGCAGAAATATGGGGGGTGAAATTAGGA 4536
          |||||||||||| ||||||||||| |||||||||||||| |||||||||| || || | ||
Db   4173 ATCTCAAAAATCTCAGGTTCACAACTCGGCAAAACAGCAGCAAAAATGTGGGCGGCTGAAATGGGGC 4232

Qy   4537 TATCGGTTGGTAAAAATCAACATAATTTTAT 4566
          |||||| |||||||||||||||||||||||
Db   4233 TATCGTTGGGTAAAAATCAACATAATTTTAT 4262
```

RESULT 3
AA211252
ID   AA211252 standard; DNA; 4350 BP.
XX
AC   AA211252;
XX
DT   08-OCT-1999 (first entry)
XX
DE   Neisseria meningitidis strain A complete ORF1 sequence.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhea; ss.
XX
OS   Neisseria meningitidis.
XX
PN   WO9924578-A2.
XX
PD   20-MAY-1999.
XX
PF   09-OCT-1998; 98MO-IB01665.
XX
PR   01-SEP-1998; 96GB-0019016.
PR   06-NOV-1997; 97GB-0023516.
PR   14-NOV-1997; 97GB-0024190.
PR   18-NOV-1997; 97GB-0024386.
PR   27-NOV-1997; 97GB-0025158.
PR   10-DEC-1997; 97GB-0026147.
PR   14-JAN-1998; 98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;
XX
DR   WPI; 1999-327407/27.
DR   P-PSDB; AAY38824.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 9; Page 364-365; 524pp; English.
XX
CC   Nucleotide sequences AA211972-Z12358 represent open reading frames
CC   (ORFs) of Neisseria meningitidis and N. gonorrhoeae which encode
CC   antigenic proteins (see AAY38499-Y38944). The antigenic proteins, their
CC   fragments, their nucleic acids and antibodies are used for diagnosis,
CC   prevention (as vaccines) or treatment of Neisseria infections,
CC   such as meningitis, septicaemia and gonorrhea. Both organisms
CC   are closely related. Fragments of the nucleic acids are useful
CC   as hybridisation probes and antisense reagents.
XX
SQ   Sequence 4350 BP; 1225 A; 1161 C; 1060 G; 842 T; 62 other;

Query Match   24.4%; Score 1177.6; DB 20; Length 4350;
Best Local Similarity   57.2%; Pred. No. 3.8e-253;
Matches 2512; Conservative 0; Mismatches 1578; Indels 303; Gaps 12;
```

```
Qy    339 TTTTTAACCGCTTGCATTCATTAGGATAGTATCGGAAGCGTGGGCGCAGGTCATACTTA 398
          ||  ||| |||||||||||   ||||||||   |||||||||||||||||||| |||||
Db     78 TTACTTAGCCTATAGCCTGTCGTTCGGCCATTCTTCCCCAAGCTTGGGCGGCACACTTA 137

Qy    399 TTTTGGATTGACTTACCAATATATATTCGTGATTTTGCCGAGAATAAAGGAAGTTCACAGT 458
          ||| || ||| |||||||||| ||| | |||| ||||||||| ||||||| || ||| |||
Db    138 TTTCGGCATCAACTACCAATACTATCGCGACACTTTGCCGGAAATAAAAGGCAAGTTTGCAGT 197

Qy    459 TGGGGCTAAAAAATATTGGAGTGGGTTTACATAATAAAAATGAAAATTTAGTTGGCACATCAATGAC 518
          ||||||| ||| || || | || |  |||||||   ||||  ||||| | |||| ||| |  ||||
Db    198 CGGGGCGAAAGATATTGAGGGTGGCNTACAACAAAAAAAGGGGAGTTGGTCGGCAAAATCAATGAC 257

Qy    519 AAAAAGCCCCAAATGATTGATCTTTCCGGTGGTGTCGGCCAAAATGGGTGGGTTGGTGGCGGAAATTGGG 578
          |||| ||||| |||||||||  |||| ||||||| ||||||| ||||| | |||||||||||| |
Db    258 AAAAGCCCGGATGATTGAGTCATTGTTCGTGGTGTCGGCCAAAATGGGTGGGTTGGTCGGCAATGGGG 317

Qy    579 CGATCAGTATATATTGTTGAGTGTGGTGCACATAAAATGAAGCTATACCAAATGTGGATTTTTGGTGC 638
          ||||||| || || |||| || | |||| || | | ||||| ||| |  |||||| |||  |||||||
Db    318 CGATCAATATATTGTTGAGCGTGGCACATAAACAATTGGCACATAAACGCGGGCTATAAACAAGTTGATTTTGGTGC 377

Qy    639 TGAAGGACAAAAATCCGTGATCAACAATCGTTTACTTATAAAAATGTGAAAACGGGAATAATTA 698
          ||| ||||| ||||||  || ||||| | | ||||| | |||||||| |||| |||| |||||
Db    378 GGAAGGAAGNAAATCCCGATCAACAATCGTTTACTTATAAAAATTGTTGAAAACGGAGATAATTA 437

Qy    699 TAAAAAACGATCAAAAGCATCCTTATGAGAAAGACTACCACAACCACGCTTACATAAATT 758
          ||| ||| |   ||||    |||   || |||||||| ||||| |||||    ||  |||
Db    438 TAAGCCTGACAAATTCACACCCTTACAACGGCCGATTANCATATGCCCGTTTGCATAAATT 497

Qy    759 TGTTACGGAAGCCACCCAAATCGAATAAATGAACTTCGATGATAATGAACGCAACAAAATATACAGA 818
          ||| ||| | || | | |||||||| ||||   |||   |  ||||| | | |||| || |  | ||
Db    498 TGTCACAGATGCAGAAACCTGTCAAAATGGGACGAGTGACATGANGGGGAAACAACAAAGTGTCCGA 557

Qy    819 TAGGACGAAAATATCCGAACGCGTGCGTATTCGGGCTCCCGGGTGGCAGTTTTGGGGAAACGA 878
          |  ||  ||||||||| | ||||| | | |   || |  ||| |  ||  |  |||||| ||
Db    558 TAAAGAAAAATATCCCGACGCGTGCCCTCAGGACCACCACTATTGCGTTATGA 617

Qy    879 TCAAAACAACGGCGACCAAGTGCGGGGGCATGATATCATTAACGTGACAAGCCAATACACA 938
          ||| | |||||| |  |||||   | ||||| || | ||||||| |||| | ||| |||||
Db    618 TGATGACAAAACACGGCGAGTTATCGTCATCTGCCGTGTCAAGTCTCGAGCGTTAATTGGCGGGG 677

Qy    939 CAACCAAGGCGGAGCCAGGGGGCGGCTGGTCAAGTGCGGCGATGTGGCCAAGCGGG 998
          || |  || |||  | | ||||   ||||| |||| || ||||| ||| | |||
Db    678 TATGCAGGGTTGGGGAAATAATGGCGTANTTAGTTTGAGCGGCGAGTGCCCATCCAA 737

Qy    999 CAATTACGGACGGCGCCCATTCCTTATTCGCAGGTCAAGCGGGACGGAGCGGCCTATGTTTAT 1058
          ||| | |||| | |  | |  |  | || |||| | |||| | |  |  | | ||| |||||
Db    738 CGACTATGGCCCTATGCCGAGTCGTTGACGGAATCTTTGACGAATGAAGAAGAAGTTTTAT 797

Qy   1059 TTATGATGGCGGAAAACGGCGGGGGGAGTTGAGGGGTGGCCAACCCTTG 1118
          ||| || ||| |||| ||   | || || | |||  || | ||| | |
Db    798 TTATGACAAAACAACAAATAAATGGCTCTCCAACGGCGAGTTTAGCCTTA 857

Qy   1119 GGGGGGGACAGAGAGAATACATTCCAACTGGTACCCCAGTCTCTTTTTTTTGATGAAAATCCTGA 1178
          ||| | ||| | | |  | | ||| | || ||| | | |  |  | |  ||| || | |||||
Db    858 TTCCGGCAGGGAAAACGGTTTCCAGCTGATACGGCCGTTNTCCAAGCACAGACATTTA 917

Qy   1179 AAAA------GATTTGCCTACATCGTTTTATAGCCCATCGGGCAATGGTCATCACAACATTAC 1235
          ||||       ||| || |  ||| | | |  ||||  ||| |  |||| | | |||||||||
Db    858 CAGAGCGGAATACACACTACCGTCTNTTTGAAACGCGGAGTAACGGACACTTTTCCTTTAC 977

Qy   1236 AGACAAAAGGCGACGCCAGGCGGCCATTGT----CAAACAACAACGAAGGACCACTGAA 1290
          | || |||| | |  | || | |  | |     ||| ||| |   | || ||  ||||
Db    918 CAGAGCGGAATACACACTACCGTCTNTTTGAAACGCGGAGTAACGGACACTTTTCCTTTAC 977

Qy   1291 GTCCGCA------------TCGTTTAAAAGACGACAAAATAACCTGCCGAAGGTAAAGACGA 1340
          ||| |||             |||| || |   ||  |||| | ||| | || || || |
Db    978 ATCCAACAACAACGGTACGGTACGGTACGGTAACGGAAAAAGGTTNCCAAATCCCAAA 1037

Qy   1341 TGTTTACCAATACCAAGGTCCAAAATATAATACCTGCCCTCGTTTTGAATTAACCGTGGGAAACCT 1400
          | ||||||| ||  |||| |  |||||| |||||| | | |   || || |  ||| || ||||
Db   1038 GCTTAAAGTACAGACACGTCCGACTGTTTTGACGAATCTTTGACGAATGAAGAAGAAGAACC 1097

Qy   1401 GTATTCGGAGATCAAAAAAAACGCCACTCTTATCACTTTACCTTATACCAACANCACCAAGGTGC 1460
          | || ||||| |||| |||| | |   |  | |   |  | | | || ||||  ||||  | |
Db   1098 AGTTTACGGCGCAGGGGGTGTTTAATCAGTACACGGTTAAACAGGTTAAAAACCT 1157

Qy   1461 GGGGCGTTTTGTATTTGAGGGTAACTTTACGTTATTCAGAGAAAATAATGCAACTTGGGCA 1520
          | | ||||| | | || ||| | |  | |  |||| || |  ||||  |  || |||| |
Db   1158 TTCGTTTTATCGATTACGGCGAACCGCAAGCATCAGGATGCTCTGAAGAAATCACCAAGCGC 1217

Qy   1521 AGGTGCTGGGAGTGCATGTAGGTGCGGAAGACAGTAGTACTGGTTACTTGGAAAAGTGTTGA 1580
          |   ||  ||  | || | | | | | |||||  | | ||| | |||  |||| | |   ||
Db   1218 GGGCGGTTTTGTATTTTGAAGGTGATTTTTACGGTTACGGTTCGCCGTTGAAAAACAACGAAAACGGTTGGGCA 1277
```